(19) Europäisches Patentamt
Europ an Patent Offic
Offi europé n d s br vets

(11) EP 0 365 657 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
04.08.1999 Bulletin 1999/31

(21) Application number: **89905317.7**

(22) Date of filing: **13.04.1989**

(51) Int Cl.[6]: **C07K 14/435**, C07K 14/72, C07H 17/02, C12N 15/00

(86) International application number:
PCT/US89/01548

(87) International publication number:
WO 89/09791 (19.10.1989 Gazette 1989/25)

(54) DNA ENCODING ANDROGEN RECEPTOR PROTEIN

FÜR ANDROGEN-REZEPTOR-PROTEIN KODIERENDE DNA

ADN CODANT POUR DES PROTEINES RECEPTRICES D'ANDROGENE

(84) Designated Contracting States:
AT BE CH DE FR GB IT LI LU NL SE

(30) Priority: **14.04.1988 US 182646**

(43) Date of publication of application:
02.05.1990 Bulletin 1990/18

(73) Proprietor: **THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL**
Chapel Hill, NC 27599-4100 (US)

(72) Inventors:
- **FRENCH, Frank, S.**
  Chapel Hill, NC 27514 (US)
- **WILSON, Elisabeth, M.**
  Chapel Hill, NC 27514 (US)
- **JOSEPH, David, R.**
  Chapel Hill, NC 27514 (US)
- **LUBAHN, Dennis, B.**
  Durham, NC 27713 (US)

(74) Representative: **MacGregor, Gordon et al**
Eric Potter Clarkson,
Park View House,
58 The Ropewalk
Nottingham NG1 5DD (GB)

(56) References cited:
WO-A-89/09223

- **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 85, October 1988, pages 7211-7215; C. CHANG et al.: "Structural analysis of complementary DNA and amino acid sequences of human and rat androgen receptors"**
- **PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE USA, vol. 86, no. 1, January 1989, pages 327-331, Washington, DC, US; W.D. TILLEY et al.: "Characterization and expression of cDNA encoding the human androgen receptor"**
- **MOLECULAR ENDOCRINOLOGY, vol. 2, no. 12, December 1988, pages 1265-1275; D.B. LUBAHN et al.: "The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate"**
- **BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS, vol. 104, no. 4, 1982, pages 1279-1286; J.A. FOEKENS et al.: "Purification of the androgen receptor of sheep seminal vesicles"**
- **Biochemical and Biophysical Research Communications, Volume 153 issued 31 May 1988, TRAPMAN "Cloning Structure and Expression of a cDNA Encoding the Human Androgen Receptor" see pages 241-248, especially figures 2 and 3.**
- **CHEMICAL ABSTRACTS, Volume 109(23) issued 05 December 1988 GOVINDAN "Cloning of the Human Androgen Receptor cDNA" see page 205.**

Printed by Jouve, 75001 PARIS (FR)

(Cont. n xt page)

EP 0 365 657 B1

- Scienc , Volume 240, issued 15 APril 1988 LUBAHN "Cloning of Human Amdrogen Rec ptor Complementary DNA Localization to the X Chromosome". see pages 327-330.
- Sci nc , Volum 240, is ued 15 April 1988 CHANG "Molecular Cloning of Human and Rat Complementary DNA Encoding Androgen Receptors" see pages 324-326.
- Nature, Volume 324, issued 18/25 December 1986 WEIMBERGER "The c- rb-A G ne Encod s a Thyroid Hormone Receptor" see pages 641-646. especially first paragraph page 641.
- Natur , Volum 320 issued 13 March 1986 GREEN "Human Oestrogen Receptor cDNA. Sequence, Expression and Homology to v-erb-A.

Description

[0001] This invention was made in the course of research supported in part by grants from the National Institutes of Health (NIH HD 16910, HD 04466, and HD 18968).

TECHNICAL FIELD OF THE INVENTION

[0002] This invention relates to recombinant DNA molecules and their expression products. More specifically this invention relates to recombinant DNA molecules coding for androgen receptor protein, androgen receptor protein, and use of the DNA molecules and protein in investigatory, diagnostic and therapeutic applications.

BACKGROUND OF THE INVENTION

[0003] The naturally occurring androgenic hormones, testosterone and its 5α-reduced metabolite, dihydrotestosterone, are synthesized by the Leydig cells of the testes and circulate throughout the body where they diffuse into cells and bind to the androgen receptor protein ("AR"). Androgens, acting through their receptor, stimulate development of the male genitalia and accessory sex glands in the fetus, virilization and growth in the pubertal male, and maintenance of male virility and reproductive function in the adult. The androgen receptor, together with other steroid hormone receptors constitute a family of trans-acting transcriptional regulatory proteins that control gene transcription through interactions with specific gene sequences.

[0004] When prostate cancer is found to be confined to the prostate gland, the treatment of choice is surgical removal. However, 50 to 80% of prostate cancer patients already have metastases at the time of diagnosis. Most of their tumors (70 to 80%) respond to the removal of androgen by castration or by suppression of luteinizing hormone secretion by the pituitary gland using a gonadotropin releasing hormone analogue alone or in combination with an anti-androgen. The degree and duration of response to this treatment is highly variable (10% live < 6 months, 50% live < 3 years, and 10% live > 10 years.) Initially cancer cells regress without androgen stimulation, but ultimately the growth of androgen independent tumor cells continues (35). At present it is not possible to predict on an individual basis which patient will respond to hormonal therapy and for how long. If poorly responsive patients could be identified early, they could be treated by alternative forms of therapy (e.g. chemotherapy) at an earlier stage when they might be more likely to respond.

[0005] Studies on androgen receptors in prostate cancer have suggested that a positive correlation may exist between the presence of androgen receptors in cancer cells and their dependence on androgenic hormone stimulation for growth. (An analogous situation exists in mammary carcinoma where there is a correlation between estrogen receptors and regression of the tumor in response to estrogen withdrawal). However, methodological problems in the measurement of androgen receptors have prevented the routine use of androgen receptor assays in the diagnostic evaluation of prostate cancer. Prior to our preparation of androgen receptor antibodies, all androgen receptor assays were based on the binding of [^{3}H]-labeled androgen. These assays have been unreliable in human prostate cancer tissue because of the extreme lability of the androgen binding site and the presence of unlabeled androgen in the tissue. Endogenous androgen occupies the binding site on the receptor and dissociates very slowly (t 1/2 24-48 hr at OC). A further problem is that biopsy samples are quite small, making it difficult to obtain sufficient tissue for [^{3}H]-androgen binding assays. Moreover, prostate cancer is heterogenous with respect to cell types. Thus within a single biopsy sample there is likely to be an uneven distribution of cells containing androgen receptors.

[0006] Development of the male phenotype and maturation of male reproductive function are dependent on the interaction of androgenic hormones with the androgen receptor protein and the subsequent function of the receptor as a trans-acting inducer of gene expression. It has become well established over the past twenty-five years that genetic defects of the androgen receptor result in a broad spectrum of developmental and functional abnormalities ranging from genetic males (46,XY) with female phenotype to phenotypically normal males with infertility. Isolation of the structural gene for the androgen receptor makes it possible to define the nature of these genomic defects in molecular terms. Analysis of the functional correlates of the genetic defects may lead to a better understanding of the regulation of androgen receptor gene expression and of the mechanism of androgen action in male sexual development and function.

[0007] The androgen insensitivity syndrome, known also as testicular feminization, is characterized by an inability to respond to androgen due to a defect in the androgen receptor, the protein that mediates the action of androgen within the cell. Androgen insensitivity is an inherited X-linked trait that occurs in both complete and incomplete forms. The complete form results in failure of male sex differentiation during embryogenesis and absence of virilization at puberty. The result is a 46,XY genetic male with testes and male internal ducts. The testes produce normal amounts of testosterone and Mullerian inhibiting substance. Consequently development of the uterus is inhibited as in the normal male. Because of the inability to respond to androgen, the external genitalia remain in the female phenotype with

normal clitoris and labia. A small vagina d velops from the urogenital sinus and ends in a blind pouch. At puberty feminization with breast development and female contours occur in response to testicular estrogen, however, there is no growth of sexual hair even though circulating testosterone concentrations are equal to or greater than levels in the normal male.

[0008] Incomplete forms of the androg n insensitivity syndrome include a spectrum of phenotypes resulting from varying degrees of incomplete androgen responsiveness. At one extreme, individuals have mild enlargement of the clitoris and sparse pubic hair. The opposite extreme is characterized by more complete masculinization with varying degrees of hypospadias deformity but predominantly a male phenotype. It has been reported that some adult men with severe oligospermia or azoospermia who are otherwise normal, have defects in the androgen receptor. These may include as many as 10% of infertile males.

[0009] The genetic defect eliciting this range of abnormalities is thought to be a single biochemical event at the level of the gene for the androgen receptor. The androgen receptor is a high affinity androgen binding protein that mediates the effects of testosterone and dihydrotestosterone by functioning as a trans-acting inducer of gene expression. For proper male sexual development to occur, there is a requirement for androgen and its receptor at a critical time during embryogenesis and during puberty. The majority of individuals with the androgen insensitivity syndrome have a history of affected family members, although about a third are thought to represent new mutations of this X-linked disorder. The incidence ranges from 1 in 20,000 to 60,000 male births.

[0010] In studies of families with clinical evidence of the androgen insensitivity syndrome, four main categories were recognized that range from the most severe, complete absence of receptor binding activity in a genetic male with female phenotype, to qualitatively normal receptor in affected individuals. Second in severity are affected individuals with qualitatively abnormal androgen binding by receptor present in normal levels. Examples include the failure of sodium molybdate (a reagent often used in studies on steroid receptors) to stabilize the receptor of affected individuals when molybdate is known to stabilize the wild-type receptor. Lability of the receptor under conditions that normally cause transformation has also been reported. A third group expresses a decreased amount of receptor with wild-type in vitro binding characteristics. The final grouping contains those androgen insensitivity patients in whom no abnormality in receptor is detected. In a recent study of this form of the syndrome, the androgen receptor was as capable of binding oligonucleotides as the wild-type receptor. Indeed, with the techniques available until only recently, it has been difficult in certain cases to document an androgen receptor defect in affected individuals.

[0011] Experimental methods used in assessing receptor defects in the past have relied on the ability of receptor to bind androgen with high affinity. The limitation of this methodology is that it is not possible to distinguish between the lack of expression of the receptor and loss of androgen binding activity. An example of how inadequate methodology complicates diagnosis is the absence of detectable receptor binding activity in patients who are partially virilized. It is theoretically possible for a mutation to occur which allows the receptor with defective androgen binding activity to induce gene transcription. Biologically active truncated forms of the glucocorticoid receptor that lack steroid binding activity but retain the DNA binding domain have been demonstrated using genetically engineered mutants.

[0012] Purification of the androgen receptor has been difficult to accomplish due to its low concentration and high degree of instability. Reported attempts at purification using either conventional methods of column chromatography or steroid-affinity chromatography have yielded insufficient amounts of receptor protein to allow even the preparation of monoclonal antibodies.

[0013] An early report on the partial purification of the androgen receptor was disclosed by Mainwaring et al. in "The use of DNA - cellulose chromatography and isoelectric focusing for the characterization and partial purification of steroid-receptor complexes," Biochem.J, 134, 113-127 (1973). They used DNA-cellulose chromatography and isoelectric focusing to isolate the receptor from rat ventral prostate and determined its physiochemical properties. This group was among the first to attempt the use of steroid affinity chromatography in conjunction with conventional chromatography, using the affinity label 17B-bromoacetoxytestosterone in receptor purification (See Mainwaring et al., "Use of the affinity label 17B-bromoacetoxytestosterone in the purification of androgen receptor proteins," Perspectives in Steroid Receptor Research, (1980)). Partial purification of androgen receptor has also been attempted from other tissue sources, such as ram seminal vesicles (See Foekens et al., Molecular Cellular Endocr, 23, 173-186 (1981) and Foekens et al., "Purification of the androgen receptor of sheep seminal vesicles," Biochem Biophys Res Comm, 104, 1279-1286 (1982)). The partially purified receptor displayed characteristics of a proteolyzed receptor, but a purification of 2,000 fold was reported with a recovery of 33% (See Foekens et al., "Purification of the androgen receptor of sheep seminal vesicles," Biochem Biophys Res Comm, 104, 1279-1286 (1982)). Later attempts at purification continued to combine steroid affinity chromatography with conventional techniques, reportedly achieving significant purification, but recoveries too low for further analysis (See Chang et al., "Purification and characterization of androgen receptor from steer semenal vesicle," Biochemistry 21, 4102-4109 (1982), Chang et al., "Purification and characterization of the androgen receptor from rat ventral prostate," Biochemistry 22, 6170-6175 (1983) and Chang et al., "Affinity labeling of the androgen receptor in rat prostate cytosol with 17B-[(bromoacetyl)oxy]-5-alpha-androstan-3-one," Biochemistry 23, 2527-2533 (1984)). More recent studies examine the effectiveness of a variety of immobilized androgens for their ability

to bind the androgen receptor (See De Larminat et al., "Synthesis and evaluation of immobilized androgens for affinity chromatography in the purification of nuclear androgen receptor," The Prostate 5, 123-140 (1984) and Bruchovsky et al, "Chemical demonstration of nuclear androgen receptor following affinity chromatography with immobilized ligands, " The Prostate 10, 207-222 (1987)). Despite these efforts, the receptor has not been purified to homogeneity and the quantities of purified androgen rec ptor obtained have been insufficient for the production of antisera.

[0014] Clinical assays for the androgen receptor now include several methods. The most common is the binding of tritium-labeled hormone and measurement of binding using a charcoal adsorption assay. Either a natural androgen, such as dihydrotestosterone, or synthetic androgen, such as mibolerone or methyltrienolone (R1881), can be used. An advantage of the latter in human tissue is that it is not significantly metabolized and does not bind to the serum androgen binding protein, sex steroid binding globulin. A limitation of radioisotope labeling of receptor is interference caused by endogenous androgen. Although exchange assays for the androgen receptor have been described (See Carroll et al., J Steroid Biochem 21, 353-359 (1984) and Traish et al., J Steroid Biochem 23, 405-413 (1985)), their effectiveness is limited by the slow kinetics of dissociation of the endogenous receptor-bound androgen.

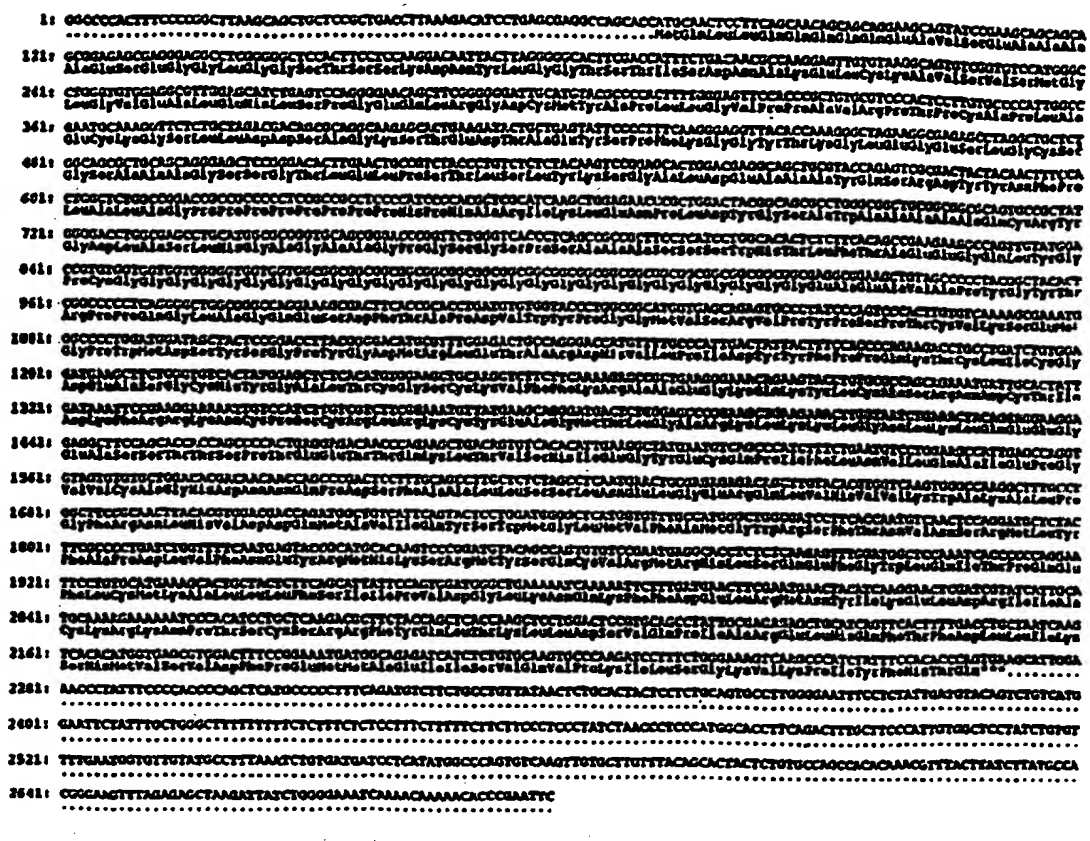
[0015] Another method used to assess receptor status is autoradiography. In this method disclosed in Barrack et al., "Current concepts and approaches to the study of prostate cancer," Progress in Clinical and Biological Research, 239, 155-187 (1987) the radioactively labeled androgen is incubated with slide-mounted tissue sections of small tissue biopsy specimens which are then frozen, sectioned and fixed. Nuclear localization of radioactivity is detected by exposure of tissue sections to x-ray film. This technique requires considerable technical expertise, is labor intensive, and requires extended periods of exposure time. It is therefore of limited usefulness in general clinical assays. Another problem is high levels of background signal, i.e. a high noise/signal ratio, making it difficult to distinguish receptor-bound nuclear radioactivity from unbound radioactivity distributed throughout the cells.

[0016] WO 87/05049 (Shine) discloses a method for the production of purified steroid receptor proteins, specifically estrogen receptor proteins, through the expression of recombinant DNA encoding for such proteins in eukaryotic host cells. However, the reference does not disclose the sequence for androgen receptor protein, nor does it disclose a method for obtaining such a sequence.

[0017] EP-A-407462 is a document which falls within the definition of Article 54(3) EPC.

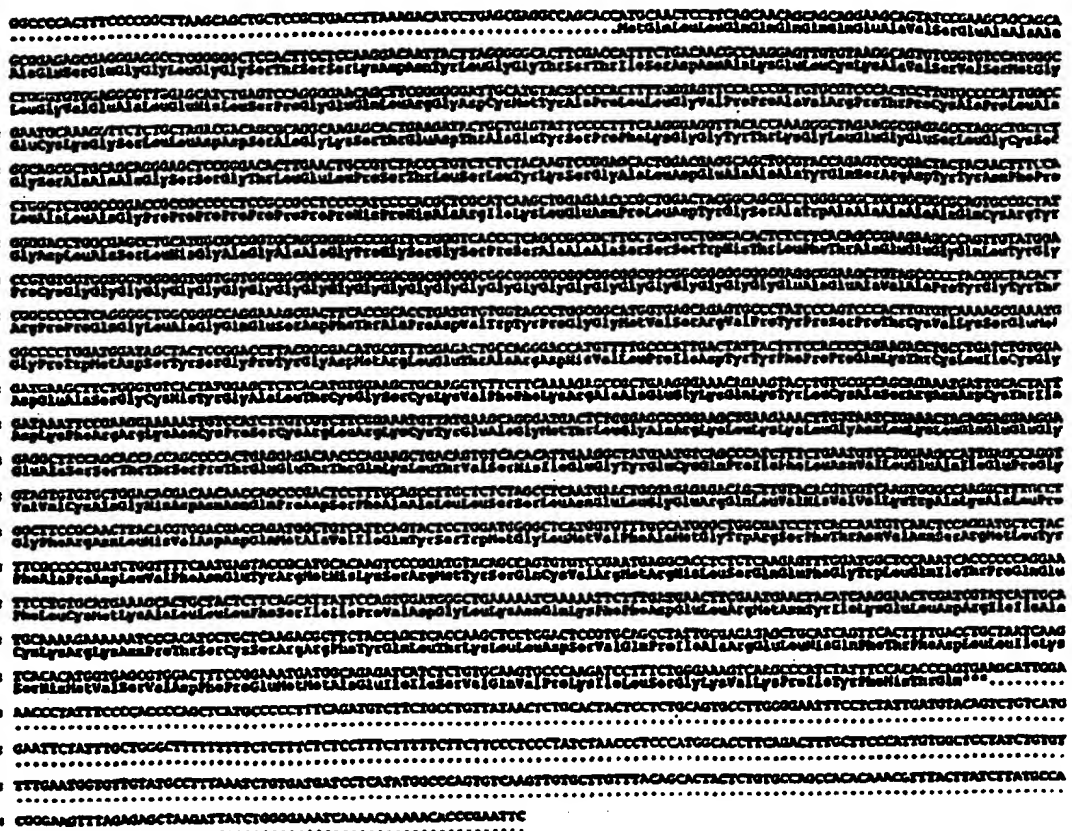
SUMMARY OF THE INVENTION

[0018] The present invention provides a recombinant DNA molecule comprising a DNA sequence that encodes for a human polypeptide which polypeptide has substantially the same biological activity as human androgen receptor protein whose amino acid sequence is shown in Figure 4 or encodes the complete amino acid sequence of Figure 4 and (i) does not hybridise under stringent conditions to a DNA molecule which codes for the following polypeptide



and (ii) is not degenerate with the said DNA molecule and (iii) is not the DNA molecule shown above. A DNA sequence encoding androgen receptor protein or a protein having substantially the same biological activity as androgen receptor activity is described. DNA sequences may be obtained from cDNA or genomic DNA, or prepared using DNA synthesis techniques.

[0019] The invention further provides a cloning vehicle comprising a DNA molecule which upon expression in a host produces a human polypeptide which polypeptide has substantially the same biological activity as human androgen receptor protein whose amino acid sequence is shown in Figure 4 or encodes the complete amino acid sequence of Figure 4 wherein the DNA molecule (i) does not hybridise under stringent conditions to a DNA molecule which codes for the following polypeptide

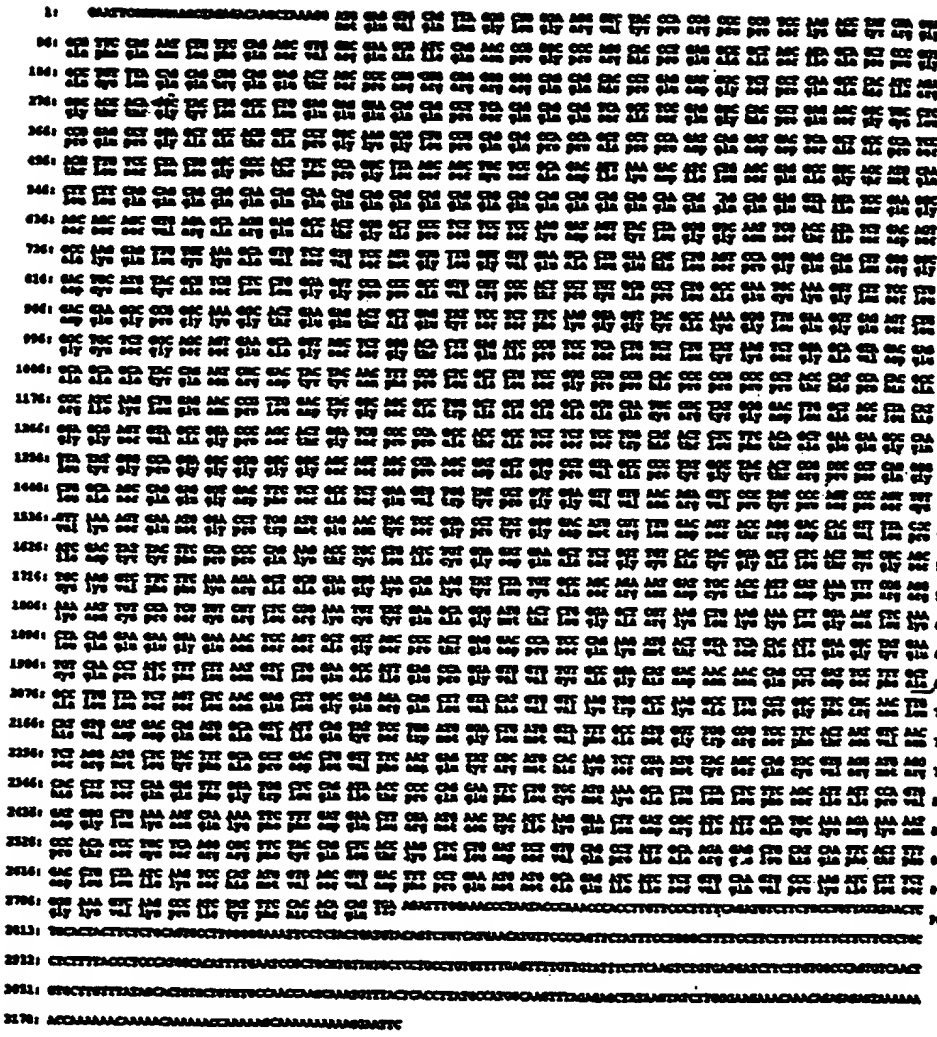


and (ii) is not degenerate with the said DNA molecule and (iii) is not the DNA molecule shown above. Cloning vehicles comprising a DNA sequence encoding androgen receptor protein or a protein having substantially the same biological activity as androgen receptor protein are described. The cloning vehicles further comprise a promoter sequence upstream of and operatively linked to the DNA sequence. In general the cloning vehicles will also contain a selectable marker, and, depending on the host cell used, may contain such elements as regulatory sequences, polyadenylation signals, enhancers and RNA splice sites.

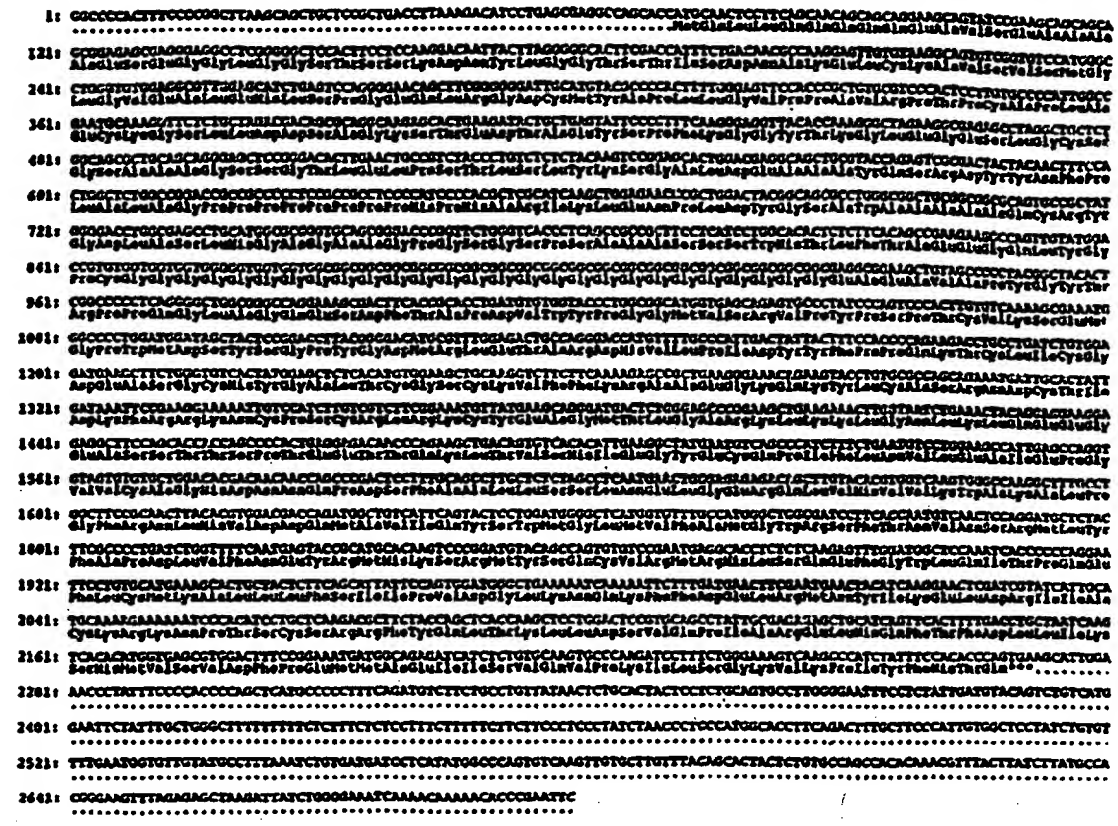
[0020] Cells transfected or transformed to produce androgen receptor protein or a protein having substantially the same biological activity as androgen receptor protein are described.

[0021] A purified androgen receptor protein and purified polypeptides and proteins having substantially the same biological activity as androgen receptor activity, and methods for producing such proteins and polypeptides are described.

[0022] The invention further provides the use of a probe comprising complementary DNA sequences derived from the deduced sequences encoding androgen receptor as shown in Figure 4 in the manufacture of a reagent to detect the presence of androgen receptor mRNA in tumour cells or to detect abnormalities in the androgen receptor gene or in its mRNA provided that the probe (i) does not hybridise under stringent conditions to a DNA molecule which codes for the polypepide



and (ii) does not hybridise under stringent conditions to DNA molecule which codes for the polypeptide



and (iii) is not degenerate with either of the said DNA molecules and (iv) is not either of the DNA molecules shown above.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] Figure 1 shows a comparison of DNA-binding domains of the human androgen receptor (hAR) with members of the nuclear receptor family. (A) is a comparison of oligo A nucleotide sequence with sequences of hAR and other nuclear receptors: hPR, human progesterone receptor; hMR, human mineralocorticoid receptor; hGR, human glucocorticoid receptor; hER, human estrogen receptor; hT3R, human thyroid hormone receptor; hRAR, human retinoic acid receptor. Chromosomal locations are shown in parentheses at the left. Nucleotide identity between oligo A and hAR is indicated with an asterisk. The percent homology with oligo A is in parentheses at the right of each sequence. (B) shows the the structure of fibroblast clone ARHFL1 human fibroblast clone [1]). Nucleotide residues are numbered from the 5'-terminus. Restriction endonuclease sites were determined by mapping or were deduced from DNA sequence. The TGA translation termination codon, determined by comparison with hPR, hMR and hGR, follows a long open reading frame containing sequences homologous to those of other steroid receptors. Arrows indicate exon boundaries in genomic clone X05AR. The hatched area is the putative DNA binding domain. (C) shows a comparison of amino acid sequences of the AR DNA-binding domain with sequences of the nuclear receptor family. AR amino acid sequence was deduced from nucleotide sequence of clone ARHFL1 and is numbered beginning with the first conserved cysteine residue (+). Amino acid numbers in parentheses at the left indicate the residue number of the first conserved cysteine from the references indicated above. Percent homology with hAR is indicated in parentheses on the right. The region of the DNA-binding domain from which the oligo A sequence was derived is underlined in hAR. Coding DNA of residues 1 to 31 is contained within genomic clone XO5AR. Abbreviations in addition to those described above are cVDR, chicken vitamin D receptor, and vERBA, erb A protein from avian erythroblastosis virus.

[0024] Abbreviations for amino acid residues are:

A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

[0025] Figure 2 illustrates the steroid binding properties of expressed AR cDNA. (A) shows the structure of pCMVAR in the expression vector pCMV containing the human cytomegalovirus(CMV) promoter of the immediate early gene, poly(A) addition-transcription terminator region of the human growth hormone gene (hGH poly A),SV40 origin of replication (SV40 Ori), and a polylinker region for insertion of cDNAs. The plasmid pTEBR contains the ampicillin resistance gene (Amp). (B) shows saturation analysis of [^{3}H]dihydrotestosterone binding in extracts of pCMVAR transfection of COS M6 cells. Portions of cytosol (0.1 ml, 0.3 mg/ml protein) were incubated overnight at 4°C with increasing concentrations of ^{3}H-labeled hormone and analyzed by charcoal adsorption. Nonspecific binding increased from 18% to 37% of total bound radioactivity. (C) shows a Scatchard plot analysis of [^{3}H]dihydrotestosterone binding. Error estimation was based on linear regression analysis (r=0.966). (U) illustrates the competition of unlabeled steroids for binding of 5 nM [^{3}H]dihydrotestosterone in transfected COS M6 cell extracts. Unlabeled steroids were added at 10- and 100-fold excess of labeled hormone. Specific binding was determined as previously described.

[0026] Figure 3 is a compiled clone map of the human androgen receptor. The map shows the structure of the human androgen receptor gene and the relative positions of the nucleic acid sequences contained in the cDNA probes [A], [B], [C] and [D], human fibroblast clone [1], human epididymis clones [1] and [5], human genomic clones [1], [2], [3], [4] and [5], and rat epididymis clones [1] and [2].

[0027] Figure 4 shows the complete nucleotide sequence for human androgen receptor cDNA and the deduced amino acid sequence.

[0028] Figure 5 shows the complete nucleotide sequence of the rat androgen receptor cDNA and the predicted amino acid sequence.

[0029] Figure 6 is a photograph of a frozen section of rat ventral prostate stained with antibodies (AR-52-3-p) to the AR peptide NH2-Asp-His-Val-Leu-Pro-Ile-Asp-Tyr-Tyr-Phe-Pro-Pro-Gln-Lys-Thr in a dilution of 1 to 3000 using the avidin-biotin peroxidase technique. Androgen receptor is indicated by brown staining of nuclei in epithelial cells.

[0030] Figure 7 is a photograph showing restriction fragment length polymorphisms in the human androgen receptor gene.

[0031] Figure 8 is a photograph showing a Southern blot analysis in the human androgen receptor gene in complete androgen insensitivity syndrome patients.

DETAILED DESCRIPTION OF THE INVENTION

[0032] In the description the following terms are employed:

Nucleotide

[0033] A monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate,, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is a nucleotide. The base characterizes the nucleotide. The four DNA bases are adenine ("A"), guanine ("G"), cytosine ("C") and thymine ("T"). The four RNA bases are A, G, C and uracil ("U"). A and G are purines, abbreviated to R, and C, T, and U are pyrimidines, abbreviated to Y.

DNA Sequence

[0034] A linear series of nucleotides connected one to the other by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

Codon

[0035] A DNA sequence of three nucleotides (a triplet) which encodes through mRNA an amino acid, a translational start signal or a translational termination signal. For example, the nucleotide triplets TTA, TTG, CTT, CTC, CTA and CTG encode for the amino acid leucine ("Leu"), TAG, TAA and TGA are translational stop signals and ATG is a translational start signal.

Reading Frame

[0036] The grouping of codons during translation of mRNA into amino acid sequences. During translation the proper reading frame must be maintained. For example, the sequence GCTGGTTGTAAG may be translated in three reading frames or phases, each of which affords a different amino acid sequence:

```
GCT GGT TGT AAG - Ala-Gly-Cys-Lys
G CTG GTT GTA AG - Leu-Val-Val
GC TGG TTG TAA A - Trp-Leu-(STOP)
Polypeptide
```

[0037] A linear series of amino acids connected one to the other by peptide bonds between the α-amino and carboxy groups of adjacent amino acids.

Genome

[0038] The entire DNA of a substance. It includes inter alia the structural genes encoding for the polypeptides of the substance, as well as operator, promoter and ribosome binding and interaction sequences including sequences such as the Shine-Dalgarno sequences.

Structural Gene

[0039] A DNA sequence which encodes through its template or messenger RNA ("mRNA") a sequence of amino acids characteristic of a specific polypeptide.

Transcription

[0040] The process of producing mRNA from a structural gene.

Translation

[0041] The process of producing a polypeptide from mRNA.

Expression

[0042] The process undergone by a structural gene to produce a polypeptide. It is a combination of transcription and translation.

Plasmid

[0043] A non-chromosomal double-stranded DNA sequence comprising an intact "replicon" such that the plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism are changed or transformed as a result of the DNA of the plasmid. For example, a plasmid carrying the gene for tetracycline resistance (Tet^R) transforms a cell previously sensitive to tetracycline into one which is resistant to it. A cell transformed by a plasmid is called a "transformant".

Phage or Bacteriophage

[0044] Bacterial virus many of which include DNA sequences encapsidated in a protein envelope or coat ("capsid"). In a unicellular organism a phage may be introduced as free DNA by a process called transfection.

Cloning Vehicle

[0045] A plasmid, phage DNA or other DNA sequences which are able to replicate in a host cell, characterized by one or a small number of endonuclease recognition sites at which such DNA sequences may be cut in a determinable

fashion without attendant loss of an essential biological function of the DNA, e.g., replication, production of coat proteins or loss of promoter or binding sites, and which contain a marker suitable for use in the identification of transformed cells, e.g., tetracycline resistance or ampicillin resistance. A cloning vehicle is often called a vector.

Cloning

[0046] The selection and propagation of a single species.

Recombinant DNA Molecule

[0047] A hybrid DNA sequence comprising at least two nucleotide sequences, the first sequence not normally being found together in nature with the second.

Expression Control Sequence

[0048] A DNA sequence of nucleotides that controls and regulates expression of structural genes when operatively linked to those genes.

[0049] To attain the objects of this invention it was necessary to determine the amino acid sequence and the DNA sequence of the structural gene encoding androgen receptor protein. One conventional approach would involve starting with the purified androgen receptor protein. However, as described above, significant amounts of the protein for such purposes have not been obtained.

[0050] An alternative approach to circumvent the overwhelming difficulty of androgen receptor protein purification is direct isolation of the DNA encoding the messenger RNA for androgen receptor protein.

[0051] Our strategy for isolating AR DNA was based on evidence that the AR gene is X-linked and that no other steroid receptor gene is located on the X chromosome. Sequence data are available from cDNAs for glucocorticoid, estrogen, progesterone, mineralocorticoid and vitamin D receptors. Comparison of the derived amino acid sequences has revealed a central region of high cysteine content which was found also in the v-erb A oncogene product recently identified as the thyroid hormone receptor. Within this 61-63 amino acid region is an arrangement of 9 cysteine residues that are absolutely conserved among steroid receptors thus far characterized. The overall homology among sequences in this conserved region ranges between 40 and 90%. We assumed that AR would resemble other members of the steroid receptor family in the conserved DNA-binding domain.

[0052] A human X chromosomal library was screened with the synthetic oligo nucleotide probe A (Oligo A sequence = $^{5'}$CTT TTG AAG AAG ACC TTA CAG CCC TCA CAG GT$^{3'}$) of Figure 1 (A) designed as a consensus sequence from the conserved sequence of the DNA-binding domain of other steroid receptors. Screening the library with the oligo A probe resulted in several recombinants whose inserts were cloned into bacteriophage M13 DNA and sequenced. One recombinant clone (Charon 35 X05AR) (human genomic clone [1]) contained a sequence similar to, yet distinct from, the DNA-binding domains of other steroid receptors. It had 84% sequence identity with oligo A, while other receptor DNAs were 78% to 91% homologous with the consensus oligonucleotide.

[0053] From the nucleotide sequence just 5' of the DNA binding domain, oligonucleotide probe B (Oligo B sequence = ^{5}GGA CCA TGT TTT GCC CAT TGA CTA TTA CTT TCC ACC CC$^{3'}$) was synthesized and used to screen bacteriophage lambda gtII cDNA libraries from human epididymis and cultured human foreskin fibroblasts. Recombinant phage (unamplified) screened with this oligonucleotide by in situ hybridization revealed one positive clone in each library. The epididymal clone (gtII ARHEL1)(human epididymis clone [1]) contained the complete DNA-binding domain and approximately 1.5 kb of upstream sequence, whereas the fibroblast clone (gtII ARHFL1)(human fibroblast clone [1]) shown in Figure 1(B) contained the DNA-binding domain and 1.5 kb of downstream sequence. The DNA-binding domains of the cDNA isolates were identical to that of the genomic exon sequence.

[0054] Transient expression in monkey kidney cells (COS M6) demonstrated that the human foreskin fibroblast cDNA fragment encodes the steroid-binding domain of hAR. A DNA fragment (ARHFLIH-X) extending 5' to 3' from the Hind III site within the putative DNA-binding domain through the stop codon (TGA) was cloned into pCMV as shown in Figure 2(A). Expression was facilitated by adding to the 5' end a consensus translation initiation sequence containing the methionine codon (ATG) in reading frame. Transfection of the recombinant construct produced a protein with high-affinity for [^{3}H]dihydrotestosterone, Figure 2(C) saturable at physiological levels of hormone. See Figure 2(B). The binding constant [K_d = 2.7 (+ 1.4) × 10^{-10} M) was nearly identical to that of native AR. The level of expressed protein, 1.3 pmol per milligram of protein, was 20 to 60 times greater than that in male reproductive tissues. Mock transfections without plasmid or transfections with plasmid DNA lacking th AR insert yielded no specific binding of dihydrotestosterone. Figure 2(D) shows steroid specificity was identical to that of native AR, with highest affinity for dihydrotestosterone and testosterone, intermediate affinity for progesterone and estradiol, and low affinity for cortisol.

[0055] Figure 3 is a clone map compiled to show the human androgen receptor gene and the nucleic acid sequences

in the cDNA clones, human genomic clones, human fibroblast clones, human epididymis clones, and rat epididymis clones. Human fibroblast clone [1] extended through the stop codon or the C-terminal end of the androgen receptor protein. To isolate and elucidate the sequence of the 5' or N-terminal end of the androgen receptor protein, we used a EcoRI/SstI fragment (EcoR1 site was from the linker) from the 5' end of human epididymis clone [1] as a probe (cDNA probe [A]), to rescreen the human X chromosomal library by standard techniques. By these techniques, human genomic clone [2] was isolated and in turn used as a probe to rescreen a human epididymis library and isolate human epididymis clone [5]. The N-terminal sequence was elucidated along with the 5' flanking sequence of the androgen receptor protein and gene. Human genomic clones [3], [4] and [5] for the sequence 3' of human genomic clone [1] were obtained using cDNA probes B [a Hind III/EcoR1 fragment] and C [an EcoR1 fragment], by screening and isolating by standard techniques.

[0056] Two rat clones, rat epididymis clones [1] and [2], were isolated from a rat epididymis cDNA library using as probes the complete human epididymis clone [1] and a EcoRI/PstI fragment, cDNA probe [D], respectively. These rat clones contained the entire protein coding sequence for the rat androgen receptor, plus flanking 5' and 3' untranslated sequences which were used to confirm the sequence of the human androgen receptor.

[0057] The complete double-stranded sequence encoding the human androgen receptor protein was determined and the deduced amino acid sequence of the human androgen receptor protein is set forth in Figure 4. The cDNA sequence and the amino acid sequence for the rat androgen receptor protein is set forth in Figure 5.

[0058] Recombinant DNA clones human fibroblast clone [1] isolated from human foreskin fibroblast cDNA gt11 expression library, human epididymis clones [1] and [5] isolated from human epididymis cDNA gt11 expression library were deposited in the American Type Culture Collection with accession numbers ATCC.# 40439, ATCC # 40442 and ATCC # 40440 respectively. Human genomic clones [1], [2], [3], [4] and [5] which were isolated from human X chromosome lambda Charon 35 library available as ATCC # 57750 have been deposited with the American Type Culture Collection with accession numbers ATCC # 40441, ATCC # 40443, ATCC # 40444, ATCC # 40445 and ATCC # 40446 respectively.

[0059] A wide variety of host-cloning vehicle combinations may be usefully employed in cloning the double stranded DNA disclosed herein. For example, useful cloning vehicles may include chromosomal, non-chromosomal and synthetic DNA sequences such as various known bacterial plasmids and wider host range plasmids such as pCMV and vectors derived from combinations of plasmids and phage DNA such as plasmids which have been modified to employ phage DNA expression control sequences. Useful hosts may include bacterial hosts, yeasts and other fungi, animal or plant hosts, such as Chinese Hamster Ovary cells (CHO, or monkey kidney cells (COS M6), and other hosts. The particular selection of host-cloning vehicle combinations may be made by those of skill in the art after due consideration of factors such as the source of the DNA- i.e. genomic or cDNA.

[0060] Cloning vehicles for use in carrying out the present invention will further comprise a promoter operably linked to the DNA sequence encoding the androgen receptor protein. In some instances it is preferred that cloning vehicles further comprise an origin of replication , as well as sequences which regulate and/or enhance expression levels, depending on the host cell selected.

[0061] Techniques for transforming hosts and expressing foreign cloned DNA in them are well known in the art (See, for example, Maniatis et al., infra). Cloning vehicles used for expressing foreign genes in bacterial hosts will generally contain a selectable marker, such as a gene for antibiotic resistance, and a promoter which functions in the host cell.

[0062] Eukaryotic microorganisms, such as the yeast Saccharomyces cerevisiae, may also be used as host cells. Cloning vehicles will generally comprise a selectable marker, such as the nutritional marker TRP, which allows selection in a host strain carrying a trpt mutation. To facilitate purification of an androgen receptor protein produced in a yeast transformant, a yeast gene encoding a secreted protein may be joined to the sequence encoding androgen receptor protein.

[0063] Higher eukaryotic cells can also serve as host cells in carrying out the present invention. Cultured mammalian cells are preferred. Cloning vehicles for use in mammalian cells will comprise a promoter capable of directing the transcription of a foreign gene introduced into a mammalian cell. Also contained in the expression vector is a polyadenylation signal, located downstream of the insertion site. The polyadenylation signal can be that of the cloned androgen receptor gene, or may be derived from a heterologous gene.

[0064] A selectable marker, such as a gene that confers a selectable phenotype, is generally introduced into the cells along with the gene of interest. Preferred selectable markers include genes that confer resistance to drugs, such as neomycin, hygromycin and methotrexate. Selectable markers may be introduced into the cell on a separate plasmid at the same time as the gene of interest, or they may be introduced on the same plasmid.

[0065] The copy marker of the integrated gene sequence can be increased through amplification by using certain selectable markers. Through selection, expression levels may be substantially increased.

[0066] Androgen receptor proteins may be purified from the host cells or cell media according to the present invention using techniques well known to those in the art. Such proteins may be utilized to produce monoclonal or polyclonal antibodies according to the techniques described below.

[0067] The techniques of this invention offer considerable advances over existing technology for measurement of androgen receptor. Utilizing proteins and peptides containing the disclosed sequences monoclonal or polyclonal antibodies can be produced for use as immunochemical reagents in immunodiagnostic assays. For example, radioimmunoassays and ELISA assays can be developed utilizing these reagents which will allow detection and quantification of androgen receptor in the presence of endogenous androgen since such androgen will not interfere with antibody binding to the receptor.

[0068] Immunocytochemistry utilizing our reagents enables determination and quantification of the cellular distribution of the androgen receptor in tumor tissues, which are often heterogenous in composition. This assay offers great potential for diagnostic evaluation of prostate cancer to determine to responsiveness to androgen withdrawal therapy.

[0069] In addition, the antibodies produced using the disclosed amino acid sequences can also be used in processes for the purification of androgen receptor protein produced by the above methods. One such purification process is disclosed in Logeat, F., et al., Biochemistry vol. 24 (1985), pp. 1029-1035, which is incorporated by reference herein.

[0070] Androgen receptor proteins and polypeptides synthesized from the deduced amino acid sequence can be used as immunogens for the preparation of antibodies to the androgen receptor. Peptides for such use range in length from about 3 to about 958 amino acids in length and are preferably from about 15 to about 30 amino acids in length. Shorter peptides may have significant sequence homology to other steroid receptor proteins and larger peptides may contain multiple antigenic determinants; these properties could result in antibodies with cross-reactivities to other steroid receptor proteins.

[0071] Peptides can be synthesized from amino acid sequences in the NH_2-terminal region, the DNA-binding domain, and the carboxyl-terminal steroid binding domain. Peptide selection will be based on hydropathic plots, selecting hydrophilic regions that are more likely exposed on the receptor surface. For diagnostic purposes preferred sequences will be selected from the NH_2-terminal region where there is the least homology with other steroid receptor proteins.

[0072] Peptides for use as immunogens can be synthesized using techniques available to one of ordinary skill in the art. For example, peptides corresponding to androgen receptor sequences can be synthesized using tBOC chemistry on a Biosearch Model 9500 peptide synthesizer. Peptide purity is assessed by high pressure liquid chromatography. Peptides can be conjugated to keyhole limpet hemocyanin through cysteine residues using the coupling agent m-maleimido-benzoyl-N-hydroxysuccinimide ester. One can also prepare resin-bound peptides utilizing the p-(oxymethyl benzamide) handle to attach the C-terminal amino acid to solid-phase resin support.

[0073] Proteins and peptides of this invention can be utilized for the production of polyclonal or monoclonal antibodies. Methods for production of such antibodies are known to those of ordinary skill in the art and may be performed without undue experimentation. One method for the production of monoclonal antibodies is described in Kohler, G., et al., "Continuous Culture of Fused Cells Secreting Antibody of Predefined Specificity," Nature, vol. 256 (1975), p. 495, which is incorporated herein by reference. Polyclonal antibodies, by way of example, can be produced by the method described below.

[0074] Peptide conjugates or resin-bound peptides can be injected into rabbits according to the procedure of Vaitukaitis et al., J Clin Endocrinol Metab, 33, 988-991 (1971) using a standard immunization schedule. Antisera titers can be determined in the ELISA assay.

[0075] For example, one androgen receptor sequence, NH_2-Asp-His-Val-Leu-Pro-Ile-Asp-Tyr-Tyr-Phe-Pro-Pro-Gln-Lys-Thr in the 5' region upstream from the DNA-binding domain, was used to raise antisera in rabbits. The antisera react selectively at a dilution of 1 to 500 with the androgen receptor both in its untransformed 8-10S form and in its 4-5S transformed form. Receptor sedimentation on sucrose gradients increases from 4 to 8-10S in the presence of antiserum at high ionic strength and from 8-10S to 11-12S at low ionic strength sucrose gradients. In the ELISA reaction against the peptide used as immunogen, reactivity was detectable at 1 to 25,000 dilution. This antiserum at a dilution of 1 to 3000 was found effective in staining nuclear androgen receptor in rat prostate and other male accessory sex glands (see Figure 6).

[0076] Our invention provides new molecular probes comprising complementary DNA sequences derived from the deduced sequences encoding the androgen receptor for diagnostic purposes. Such probes may be used to detect the presence of androgen receptor mRNA in tumor cells. Such probes may also be used for detection of androgen receptor gene defects. Androgen receptor complementary DNA sequences can be used as hybridization probes to detect abnormalities in the androyen receptor gene or in its messenger RNA.

[0077] Androgen receptor DNA sequences disclosed and complementary RNA sequences can be used to construct probes for use in DNA hybridization assays. An example of one such hybridization assay and methods for constructing probes for such assays are disclosed in U.S.Patent No. 4,683,195 to Mullis et al., U.S. Patent No. 4,683,202 to Mullis, U.S. Patent No. 4,617,261 to Sheldon , III et al., U.S. Patent No. 4,683,194 to Salki et al., and U.S. Patent No. 4,705,886 to Levenson et al., which are hereby incorporated by reference.

[0078] By example, one method for detecting gene deletion utilizes Southern blotting and hybridization. DNA can be isolated from cultured skin fibroblasts or from leukocytes obtained from blood. DNA is cut with restriction enzymes, el ctrophoresed on an agarose yel, blotted onto nitrocellulose, and hybridized with [^{32}P]-labeled androgen receptor

DNA (see Maniatis, T. et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, 1982, incorporated by reference herein).

[0079] In addition, small mutations can be detected utilizing methods known to one of ordinary skill in the art, from cultured skin fibroblasts of the affected individual. A cDNA library can been prepared using standard techniques. The androgen receptor clones can be isolated using a [^{32}P] DNA AR probe. The clones AR cDNA can then be sequenced and compared to normal AR cDNA sequences.

[0080] Alternatively genomic DNA can be isolated from blood leukocytes or cultured skin fibroblasts of the affected individual. The DNA is then subjected to restriction enzyme digestion, electrophoresis and is blotted onto nitrocellulose. Synthetic oligonucleotides can be used to bracket specific exons. Exon sequences are amplified using the polymerase chain reaction, cloned into M13 and sequenced. The sequences are compared to normal human AR DNA sequences.

[0081] Another method of identifying small mutations or deletions takes advantage of the ability of RNase A to cleave regions of single stranded RNA in RNA:DNA hybrids. Genomic DNA isolated from fibroblasts of affected individuals is hybridized with radioactive RNA probes (Promego Biotec) prepared from wild-type androgen receptor cDNA. Mismatches due to mutations would be cleaved by RNase A and result in altered sized bands relative to wild-type on denaturing polyacrylamide gels.

[0082] Restriction fragment length polymorphism (RFLP) linked to the androgen receptor gene locus may be used in prenatal diagnosis and carrier detection of androgen insensitivity. For example, the presence of RFLPs in normal individuals is first established by isolating UNA from lymphocytes of at least six females (total of 12 X chromosomes). DNA can be isolated using the proteinase K procedure and fragmented using a battery of restriction enzymes. Preferred are those enzymes that contain the dinucleotide sequence CG in their recognition sequence. Southern blots are screened with 5-10 kb androgen receptor genomic fragments which if possible lack repetitive DNA. For those regions containing repetitive elements, total human genomic DNA can be added as competitor in the hybridization reaction. Alternatively, one can subclone selected regions to yield a probe free of repetitive elements.

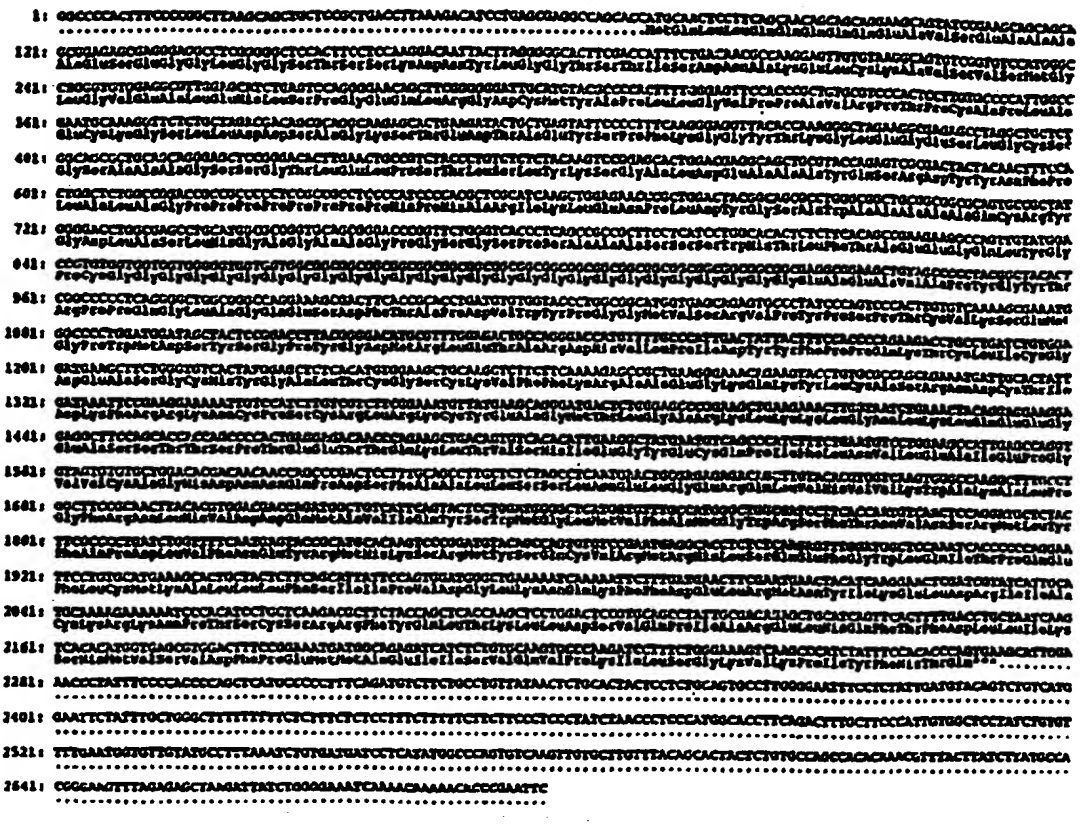
[0083] For example, a human restriction fragment length was determined by cDNA probe (B) and Hind III restriction endonuclease using the Southern blot technique (See Figure 7). The two RFLP alleles detected are a fragment at 6.5 kb (allele) and a fragment at 3.5 kb (allele 2). Major constant fragment bands are seen at approximately 2 and 5 kb with minor constant bands at 0.9 and 7.5 kb. Allele 1 is present in approximately 30% of the X chromosomes of the Caucasian population. Allele 2 is present in approximately 20% of the X chromosomes of the Caucasian population. In Figure 8 Lanes A, B and D, DNA from women who are homozygous for allele 1 is shown. In Figure 8 Lane C, DNA from a woman who is heterozygous for both alleles 1 and 2 is shown. Figure 8 Lane E contains DNA from a man that only possesses allele 2. This RFLP, and others determined by the clones we have isolated, will enable one to monitor the androgen receptor gene in various disease conditions described herein.

[0084] An example of using the androgen receptor clones to detect mutations is shown in Figure 8 where five different complete androgen insensitive patients' DNA are digested with EcoR1, electrophoresed on a Southern blot, and probed with cDNA probe B. The patient in lane B lacks a 3kb band indicating that part of the androgen receptor gene is deleted. Further analysis of this and other patients DNA is possible with other AR probes and by sequencing by standard methods and comparing the abnormal sequence to the normal sequence described herein.

[0085] Other potential uses for oligonucleotide sequences disclosed, for example in construction of therapeutics to block genetic expression, will be obvious to one of ordinary skill in the art.

Claims

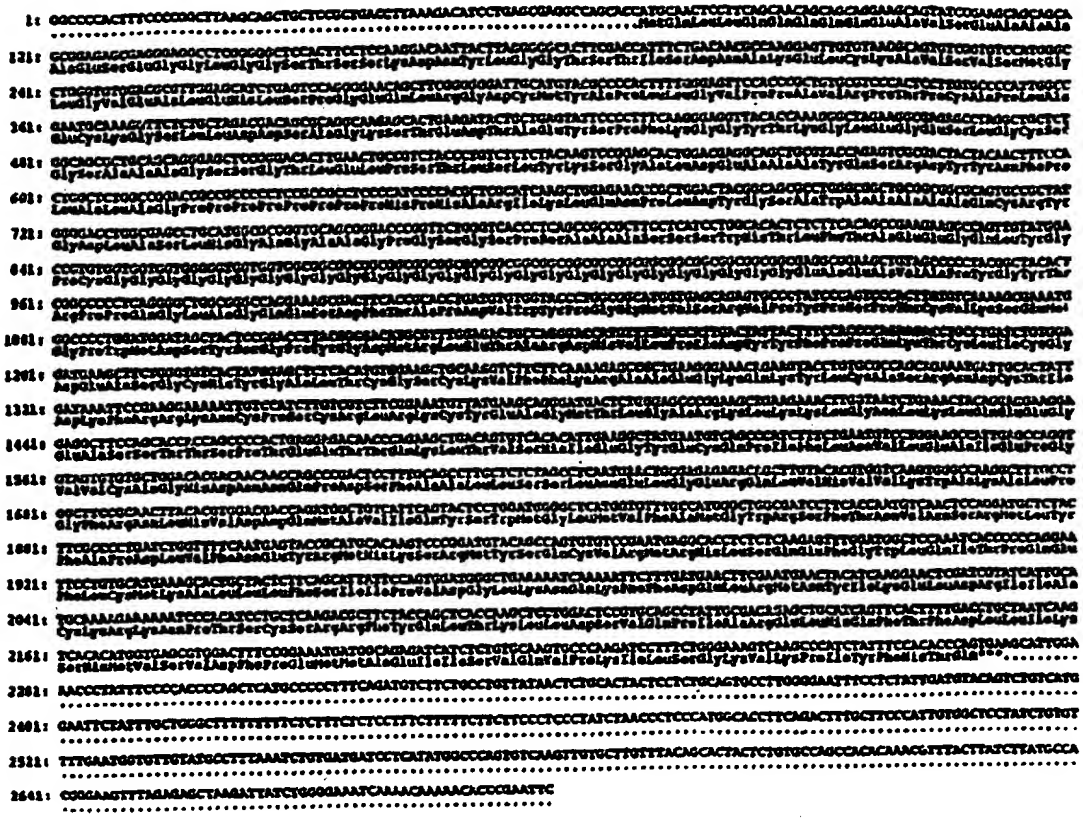
1. A recombinant DNA molecule comprising a DNA sequence that encodes for a human polypeptide which polypeptide has substantially the same biological activity as human androgen receptor protein whose amino acid sequence is shown in Figure 4 or encodes the complete amino acid sequence of Figure 4 and (i) does not hybridise under stringent conditions to a DNA molecule which codes for the following polypeptide



and (ii) is not degenerate with the said DNA molecule and (iii) is not the DNA molecule shown above.

2. The recombinant DNA molecule of Claim 1 wherein the DNA encodes the human androgen receptor protein whose amino acid sequence is shown in Figure 4.

3. A cloning vehicle comprising a DNA molecule which upon expression in a host produces a human polypeptide which polypeptide has substantially the same biological activity as human androgen receptor protein whose amino acid sequence is shown in Figure 4 or encodes the complete amino acid sequence of Figure 4 wherein the DNA molecule (i) does not hybridise under stringent conditions to a DNA molecule which codes for the following polypeptide



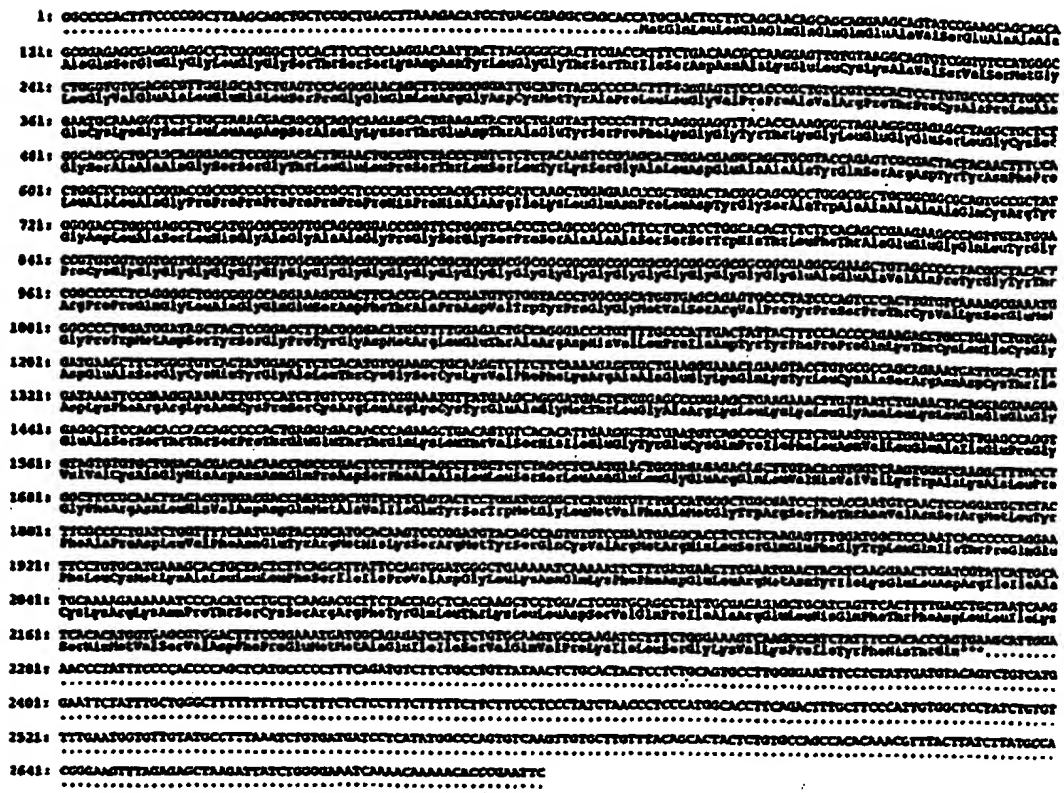
and (ii) is not degenerate with the said DNA molecule and (iii) is not the DNA molecule shown above.

4. The cloning vehicle of Claim 3 wherein the DNA molecule encodes the human androgen receptor protein whose amino acid sequence is shown in Figure 4.

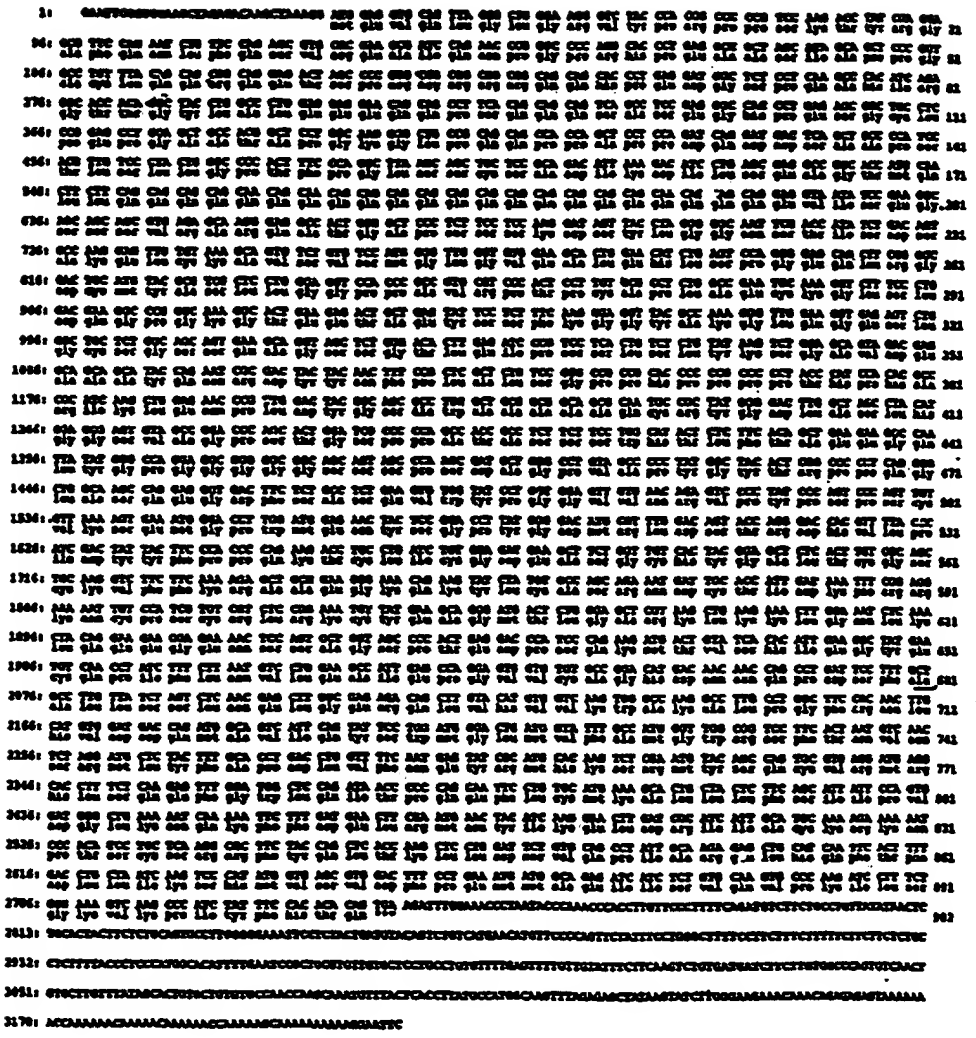
5. A process for producing an androgen receptor protein, the process comprising transfecting or transforming a host organism with the cloning vehicle of Claim 3, and purifying androgen receptor protein produced by translation of the DNA sequence encoding the protein.

6. A process for producing a human androgen receptor protein, the process comprising transfecting or transforming a host organism with the cloning vehicle of Claim 4, and purifying human androgen receptor protein produced by translation of the DNA sequence encoding the protein.

7. Use of a probe comprising complementary DNA sequences derived from the deduced sequences encoding androgen receptor as shown in Figure 4 in the manufacture of a reagent to detect the presence of androgen receptor mRNA in tumour cells or to detect abnormalities in the androgen receptor gene or in its mRNA provided that the probe (i) does not hybridise under stringent conditions to a DNA molecule which codes for the polypeptide



and (ii) does not hybridise under stringent conditions to DNA molecule which codes for the polypeptide



and (iii) is not degenerate with either of the said DNA molecules and (iv) is not either of the DNA molecules shown above.

8. A recombinant DNA molecule comprising a DNA sequence having the structural gene which encodes for human androgen receptor protein whose amino acid sequence is shown in Figure 4.

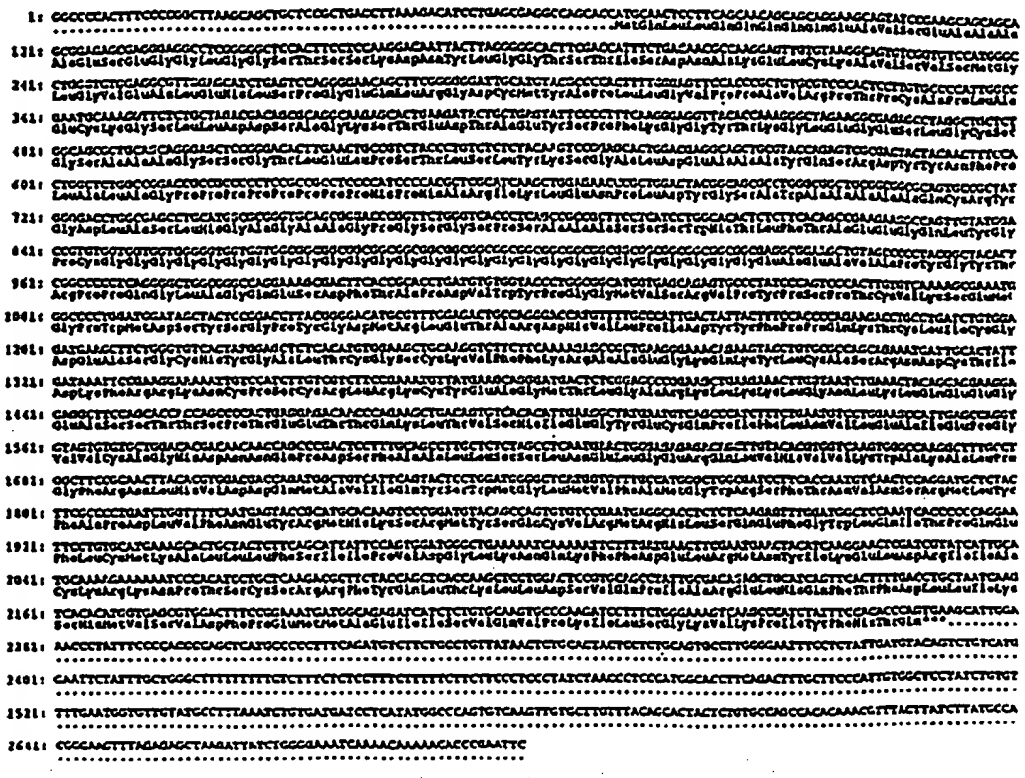
9. A cloning vehicle comprising a DNA molecule which upon expression in a host produces human androgen receptor protein whose amino acid sequence is shown in Figure 4.

10. A process for producing a human androgen receptor protein, the process comprising transfecting or transforming a host organism with the cloning vehicle of Claim 9, and purifying human androgen receptor protein produced by translation of the DNA sequence encoding the protein.

Patentansprüche

1. Ein rekombinantes DNA-Molekül mit iner DNA-Sequenz, die für ein humanes Polypeptid codiert, das im wesentlichen die gleiche biologische Aktivität wie das humane Androgen-Rezeptor-Protein aufweist, dessen Aminosäuresequenz in Figur 4 gezeigt ist, oder die für die vollständige Aminosäuresequenz von Figur 4 codiert, und das (i)

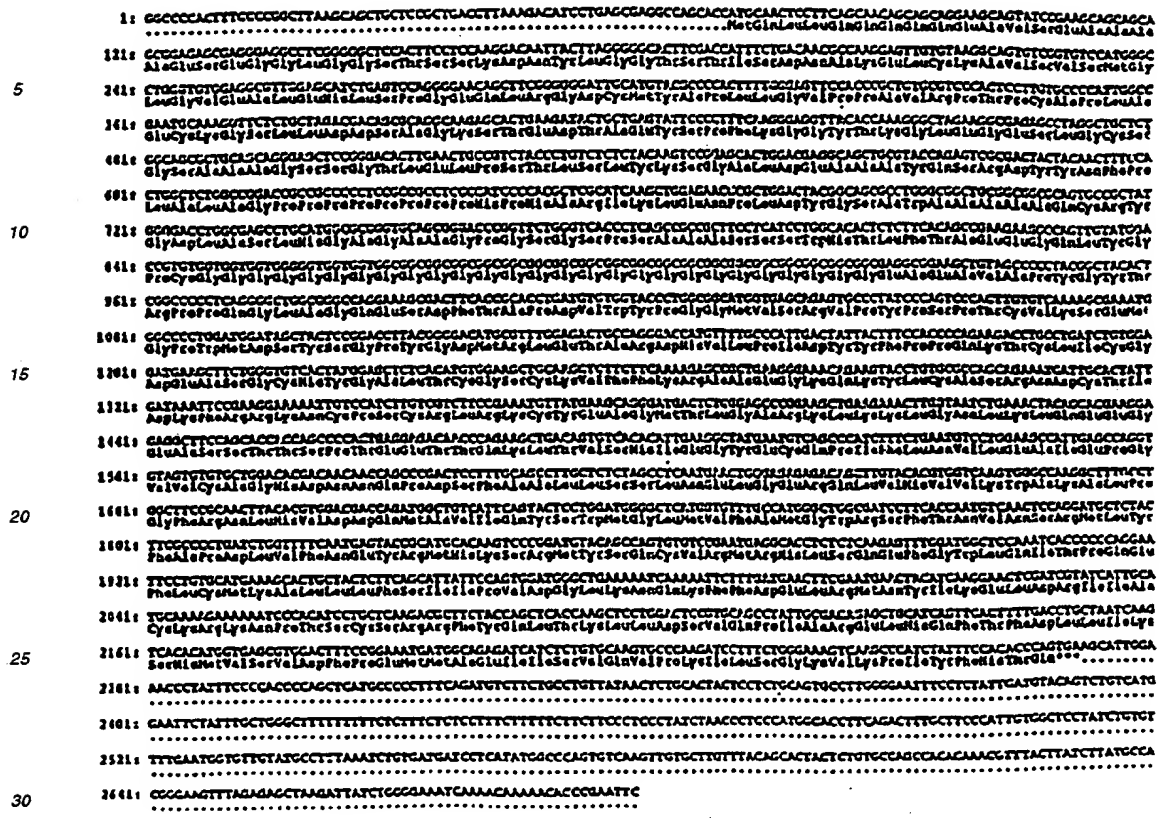
unter stringenten Bedingungen nicht mit einem DNA-Molekül hybridisiert, das für das folgende Polypeptid codiert:



und das (ii) bezüglich des genannten DNA-Moleküls nicht degeneriert ist und das (iii) nicht das oben gezeigte DNA-Molekül ist.

2. Rekombinantes DNA-Molekül nach Anspruch 1, wobei die DNA für das humane Androgen-Rezeptor-Protein codiert, dessen Aminosäuresequenz in Figur 4 gezeigt ist.

3. Klonierungsvektor mit einem DNA-Molekül, das bei der Expression in einem Wirt ein humanes Polypeptid erzeugt, wobei das Polypeptid im wesentlichen die gleiche biologische Aktivität wie das humane Androgen-Rezeptor-Protein aufweist, dessen Aminosäuresequenz in Figur 4 gezeigt ist, oder das für die vollständige Aminosäuresequenz von Figur 4 codiert, wobei das DNA-Molekül (i) unter stringenten Bedingungen nicht mit einem DNA-Molekül hybridisiert, das für das folgende Polypeptid codiert



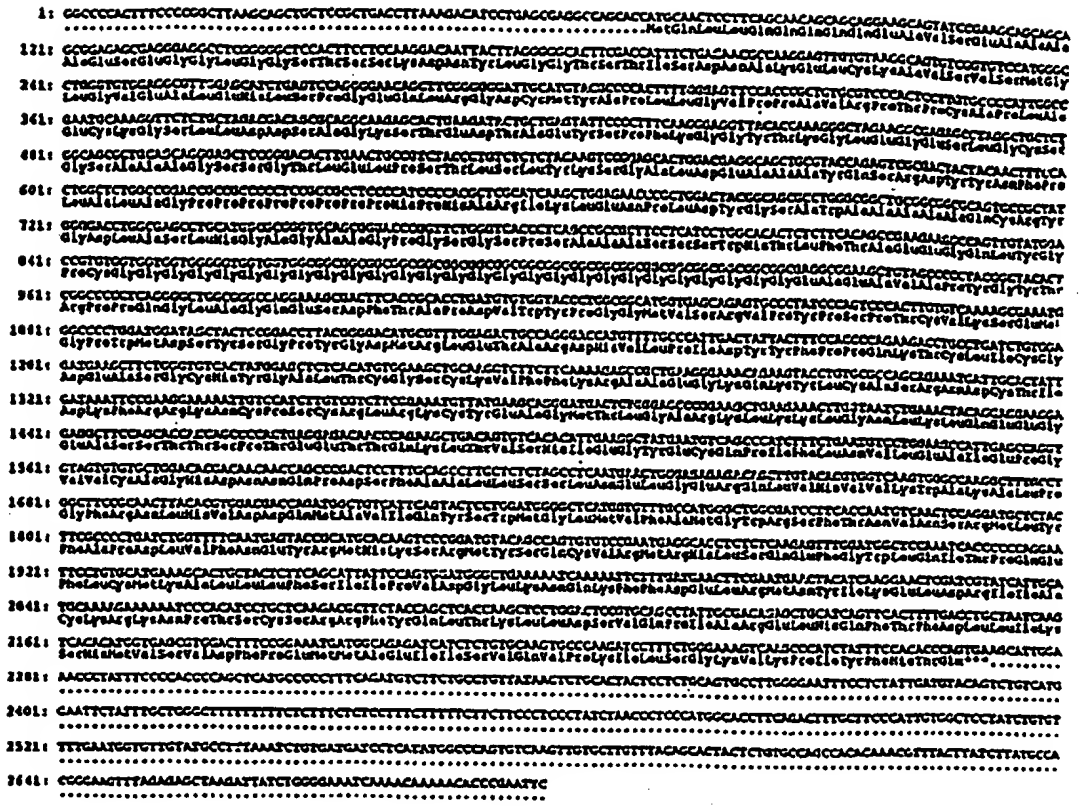
und (ii) bezüglich des genannten DNA-Moleküls nicht degeneriert ist und (iii) nicht das oben gezeigte DNA-Molekül ist.

4. Klonierungsvektor von Anspruch 3, bei dem das DNA-Molekül für das humane Androgen-Rezeptor-Protein codiert, dessen Aminosäuresequenz in Figur 4 gezeigt ist.

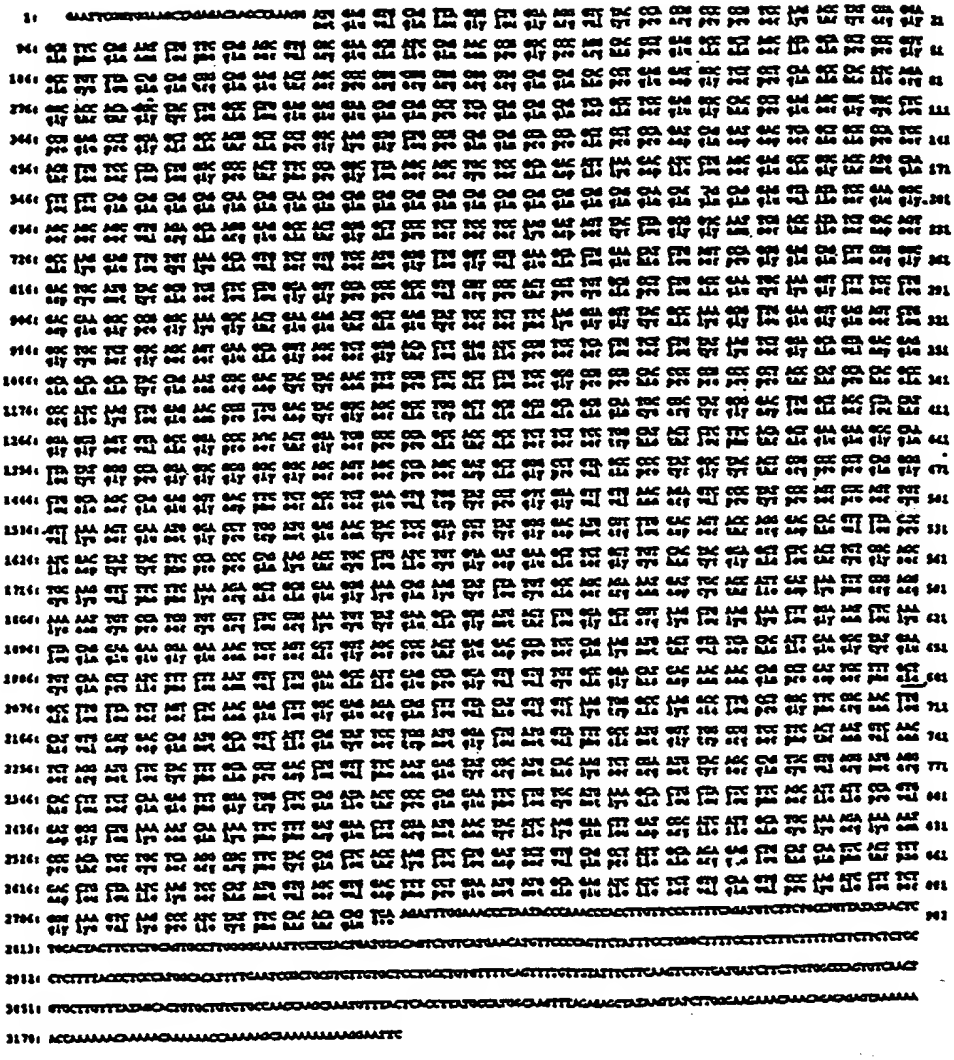
5. Verfahren zur Herstellung eines Androgen-Rezeptor-Proteins, wobei das Verfahren die Transfektion oder Transformation eines Wirtsorganismus mit dem Klonierungsvektor nach Anspruch 3 sowie die Reinigung des Androgen-Rezeptor-Proteins, das durch Translation der für das Protein codierenden DNA-Sequenz hergestellt wurde, umfaßt.

6. Verfahren zur Herstellung eines humanen Androgen-Rezeptor-Proteins, wobei das Verfahren die Transfektion oder Transformation eines Wirtsorganismus mit dem Klonierungsvektor nach Anspruch 4 sowie die Reinigung des humanen Androgen-Rezeptor-Proteins, das durch Translation der für das Protein codierenden DNA-Sequenz hergestellt wurde, umfaßt.

7. Verwendung einer Sonde mit komplementären DNA-Sequenzen, die sich von den abgeleiteten Sequenzen ableiten, die für den Androgen-Rezeptor codieren, wie er in Figur 4 gezeigt ist, bei der Herstellung eines Reagens zum Nachweis der Gegenwart von Androgen-Rezeptor-mRNA in Tumorzellen oder zum Nachweis von Anomalien im Androgen-Rezeptor-Gen oder in seiner mRNA, mit der Maßgabe, daß die Sonde (i) unter stringenten Bedingungen nicht mit einem DNA-Molekül hybridisiert, das für das Polypeptid



codiert und (ii) unter stringenten Bedingungen nicht mit einem DNA-Molekül hybridisiert, das für das Polypeptid



codiert und (iii) bezüglich keines der genannten DNA-Moleküle degeneriert ist und (iv) keines der oben gezeigten DNA-Moleküle ist.

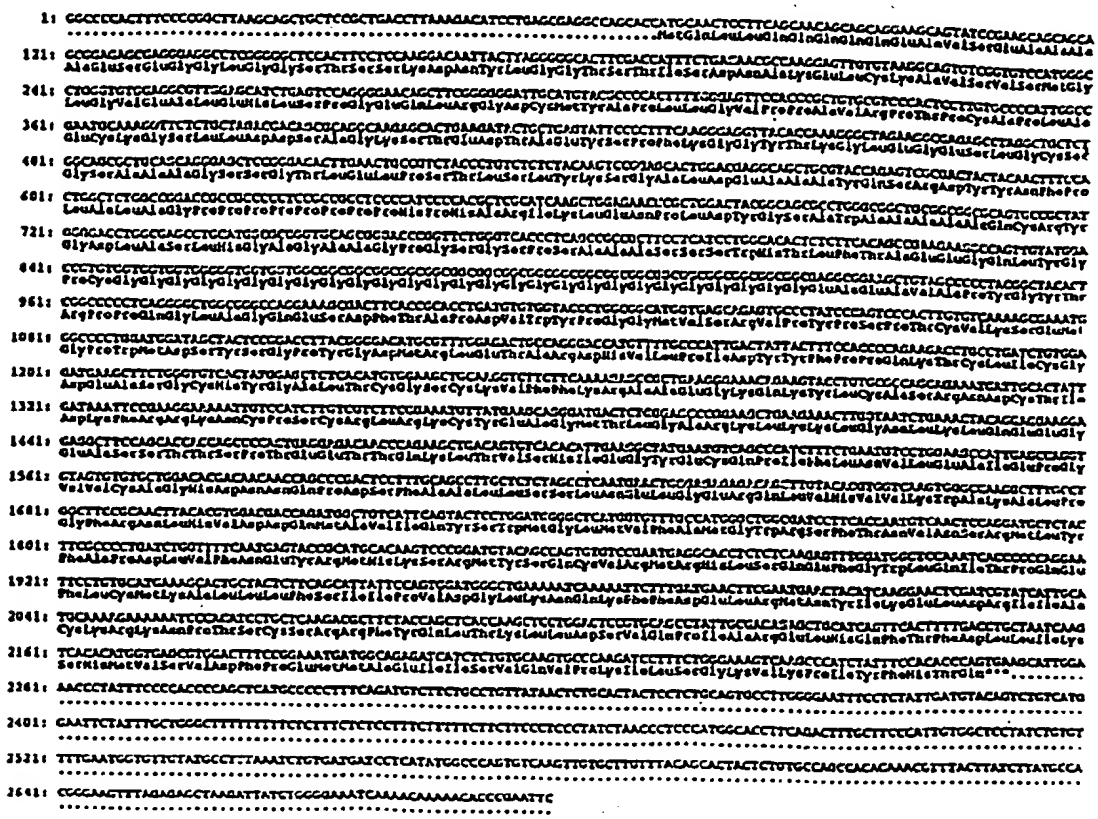
8. Rekombinantens DNA-Molekül mit einer DNA-Sequenz mit dem Strukturgen, das für das humane Androgen-Rezeptor-Protein codiert, dessen Aminosäuresequenz in Figur 4 gezeigt ist.

9. Klonierungsvektor mit einem DNA-Molekül, das bei der Expression in einem Wirt humanes Androgen-Rezeptor-Protein liefert, dessen Aminosäuresequenz in Figur 4 gezeigt ist.

10. Verfahren zur Herstellung eines humanen Androgen-Rezeptor-Proteins, wobei das Verfahren die Transfektion oder die Transformation eines Wirtsorganismus mit dem Klonierungsvektor von Anspruch 9 und die Reinigung des durch Translation der für das Protein codierenden DNA-Sequenz erhaltenen humanen Androgen-Rezeptor-Proteins umfaßt.

Revendications

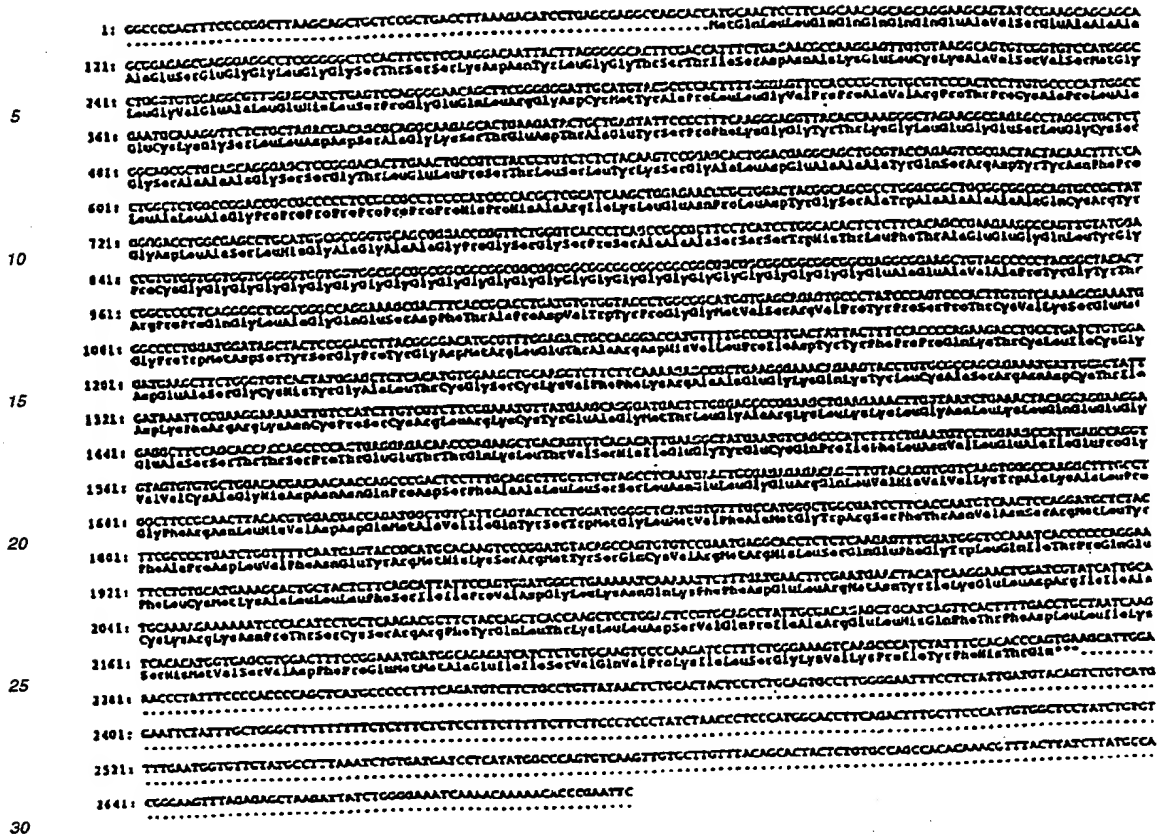
1. Molécule d'ADN recombinant comprenant une séquence d'ADN qui code pour un polypeptide humain, ce polypeptide ayant sensiblement la même activité biologique que la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4 ou qui code pour la séquence complète d'acides aminés de la figure 4 et qui (i) ne s'hybride pas sous des conditions rigoureuses à une molécule d'ADN qui code pour le polypeptide suivant :



et (ii) n'est pas dégénérée avec ladite molécule d'ADN et (iii) n'est pas la molécule d'ADN représentée ci-dessus.

2. Molécule d'ADN recombinant de la revendication 1, dans laquelle l'ADN code pour la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4.

3. Vecteur de clonage comprenant une molécule d'ADN, qui lors de l'expression dans un hôte, produit un polypeptide humain, ce polypeptide ayant sensiblement la même activité biologique que la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4 ou qui code pour la séquence complète d'acides aminés de la figure 4, dans lequel la molécule d'ADN (i) ne s'hybride pas sous des conditions rigoureuses à une molécule d'ADN qui code pour le polypeptide suivant :



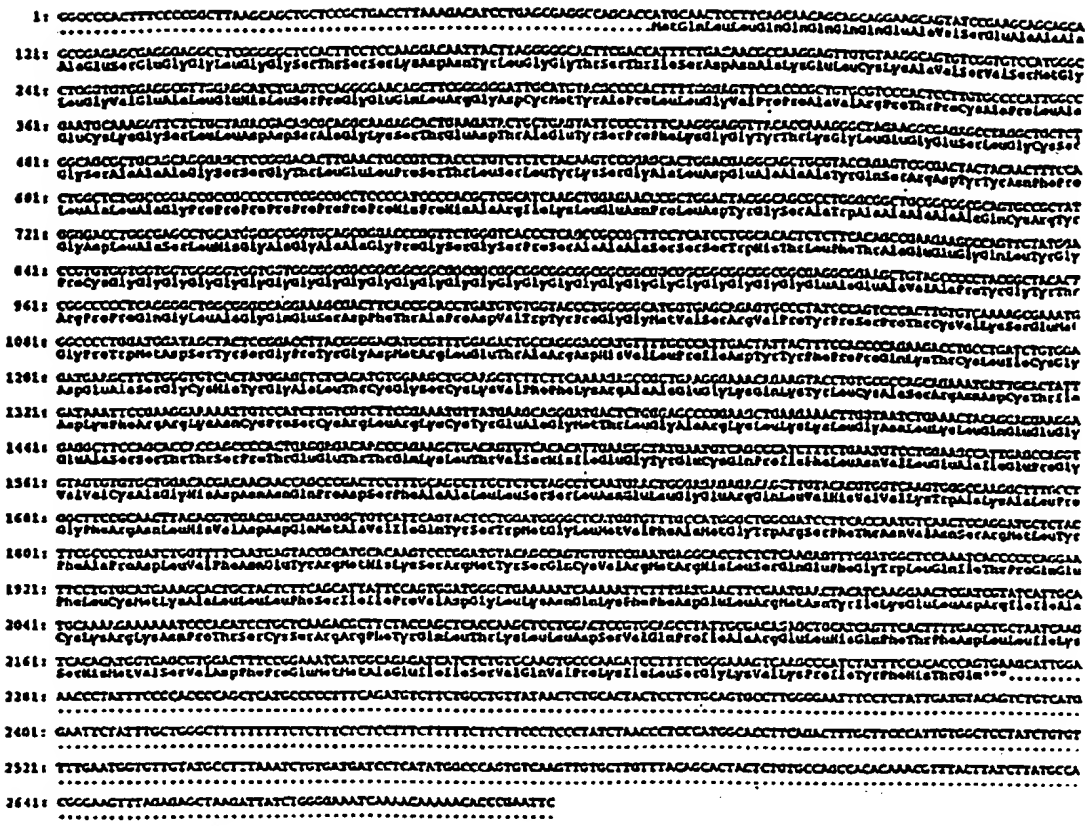
et (ii) n'est pas dégénérée avec ladite molécule d'ADN et (iii) n'est pas la molécule d'ADN représentée ci-dessus.

4. Vecteur de clonage de la revendication 3, dans lequel la molécule d'ADN code pour la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4.

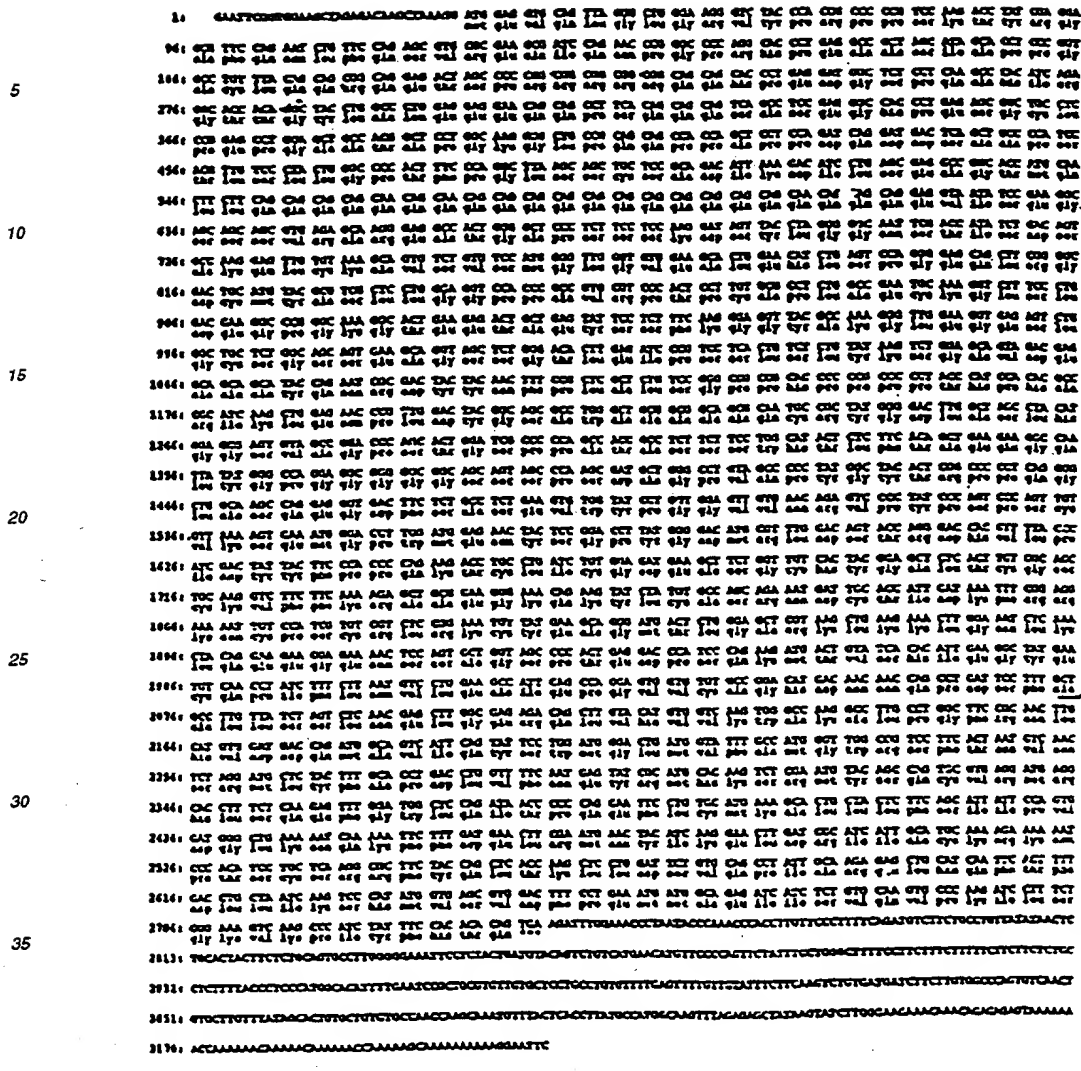
5. Procédé de production d'une protéine du récepteur androgénique, le procédé comprenant la transfection ou la transformation d'un organisme hôte avec le vecteur de clonage de la revendication 3, et la purification de la protéine du récepteur androgénique, produite par la traduction de la séquence d'ADN codant pour la protéine.

6. Procédé de production d'une protéine du récepteur androgénique humain, le procédé comprenant la transfection ou la transformation d'un organisme hôte avec le vecteur de clonage de la revendication 4, et la purification de la protéine du récepteur androgénique humain, produite par la traduction de la séquence d'ADN codant pour la protéine.

7. Utilisation d'une sonde comprenant des séquences d'ADN complémentaires dérivées des séquences déduites codant pour le récepteur androgénique comme représenté sur la figure 4, dans la fabrication d'un réactif pour détecter la présence d'ARNm du récepteur androgénique dans des cellules tumorales ou pour détecter des anomalies dans le gêne du récepteur androgénique ou dans son ARNm, pourvu que la sonde (i) ne s'hybride pas sous des conditions rigoureuses à une molécule d'ADN qui code pour le polypeptide



et (ii) ne s'hybride pas sous des conditions rigoureuses à une molécule d'ADN qui code pour le polypeptide



et (iii) n'est pas dégénérée avec l'une ou l'autre desdites molécules d'ADN et (iv) n'est ni l'une ni l'autre des molécules d'ADN représentée ci-dessus.

8. Molécule d'ADN recombinant comprenant une séquence d'ADN possédant le gêne de structure qui code pour la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4.

9. Vecteur de clonage comprenant une molécule d'ADN qui lors de l'expression dans un hôte, produit la protéine du récepteur androgénique humain dont la séquence d'acides aminés est représentée sur la figure 4.

10. Procédé de production d'une protéine du récepteur androgénique humain, le procédé comprenant la transfection ou la transformation d'un organisme hôte avec le vecteur de clonage de la revendication 9 et la purification de la protéine du récepteur androgénique humain, produite par la traduction de la séquence d'ADN codant pour la protéine.

A.

FIGURE 1 (Page 1 of 2)

Oligo A Complement	5'-ACC	TGT	GAG	GGC	TGT	AAG	GTC	TTC	TTC	AAA	AG-3'	(100%)
	**	***	*	*	**	***	***	***	***	***	**	
hAR (X)	ACA	TGT	GGA	AGC	TGC	AAG	GTC	TTC	TTC	AAA	AG	(84%)
hPR (11)	ACC	TGT	GGG	AGC	TGT	AAG	GTC	TTC	TTT	AAC	AG	(88%)
hMR (4)	ACC	TGT	GGC	AGC	TGC	AAA	GTT	TTC	TTC	AAA	AG	(81%)
hGR (5)	ACT	TGT	GGA	AGC	TGT	AAA	GTT	TTC	TTC	AAA	AG	(81%)
hER (6)	TCC	TGT	GAG	GGC	TGT	AAG	GCC	TTC	TTC	AAG	AG	(91%)
hT3R (3, 17)	ACG	TGT	GAA	GGC	TGC	AAG	GCT	TTC	TTT	AGA	AG	(78%)
hRAR (17)	GCC	TGT	GAG	GGC	TGC	AAG	GGC	TTC	TTC	CGC	CG	(78%)

B.

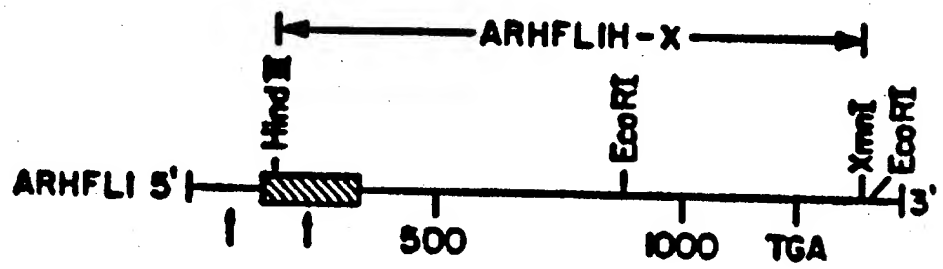


FIGURE 1 (Page 2 of 2)

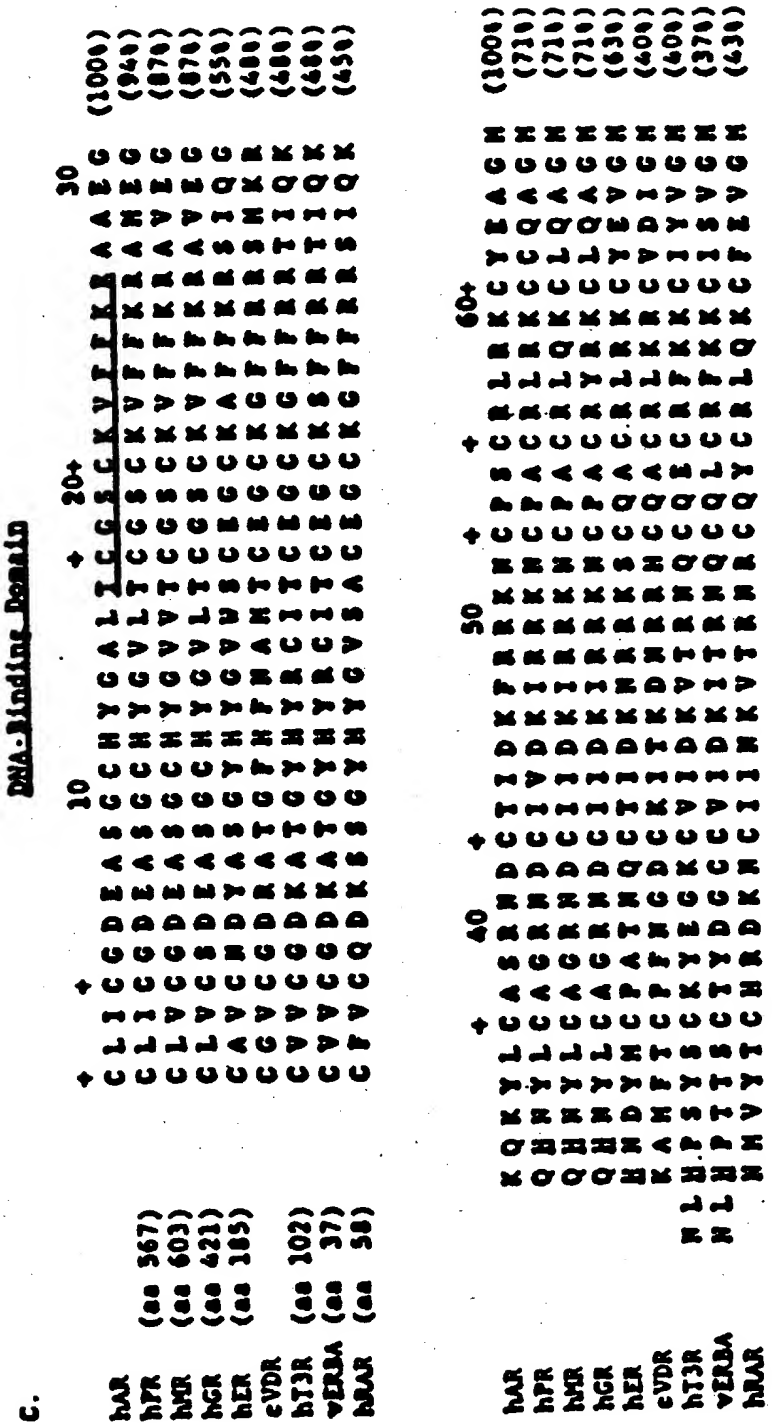
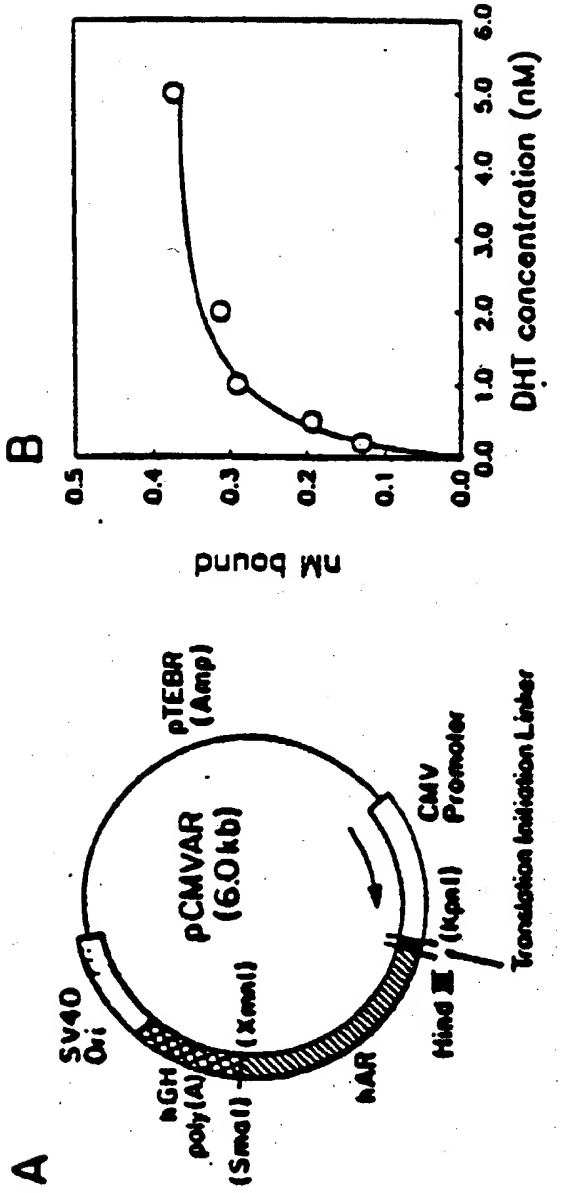


FIGURE 2 (Page 1 of 2)



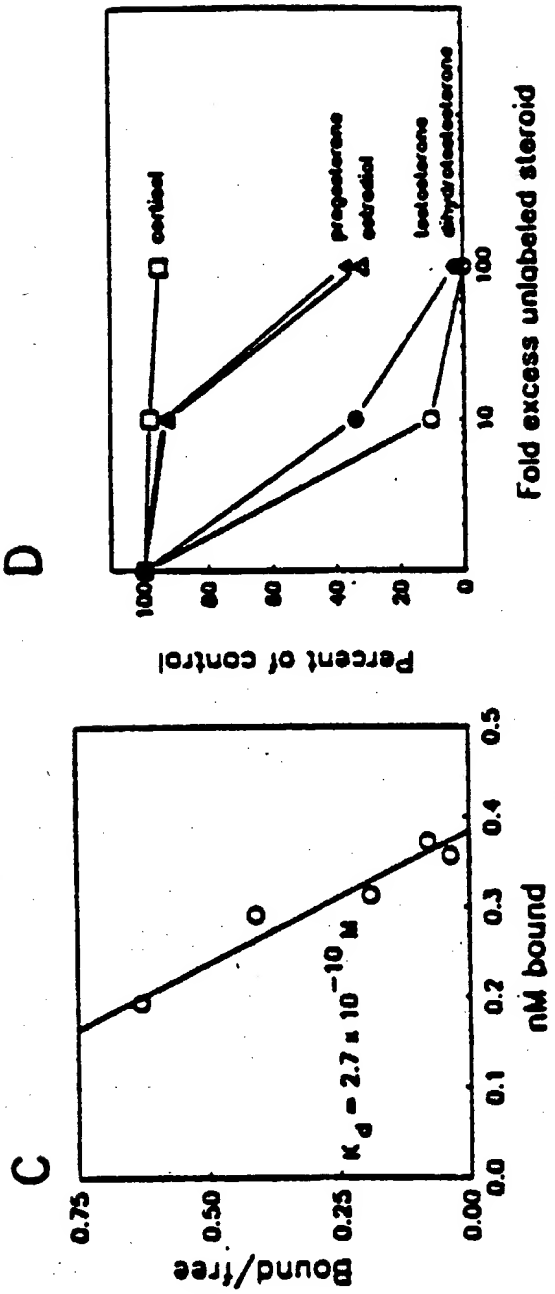
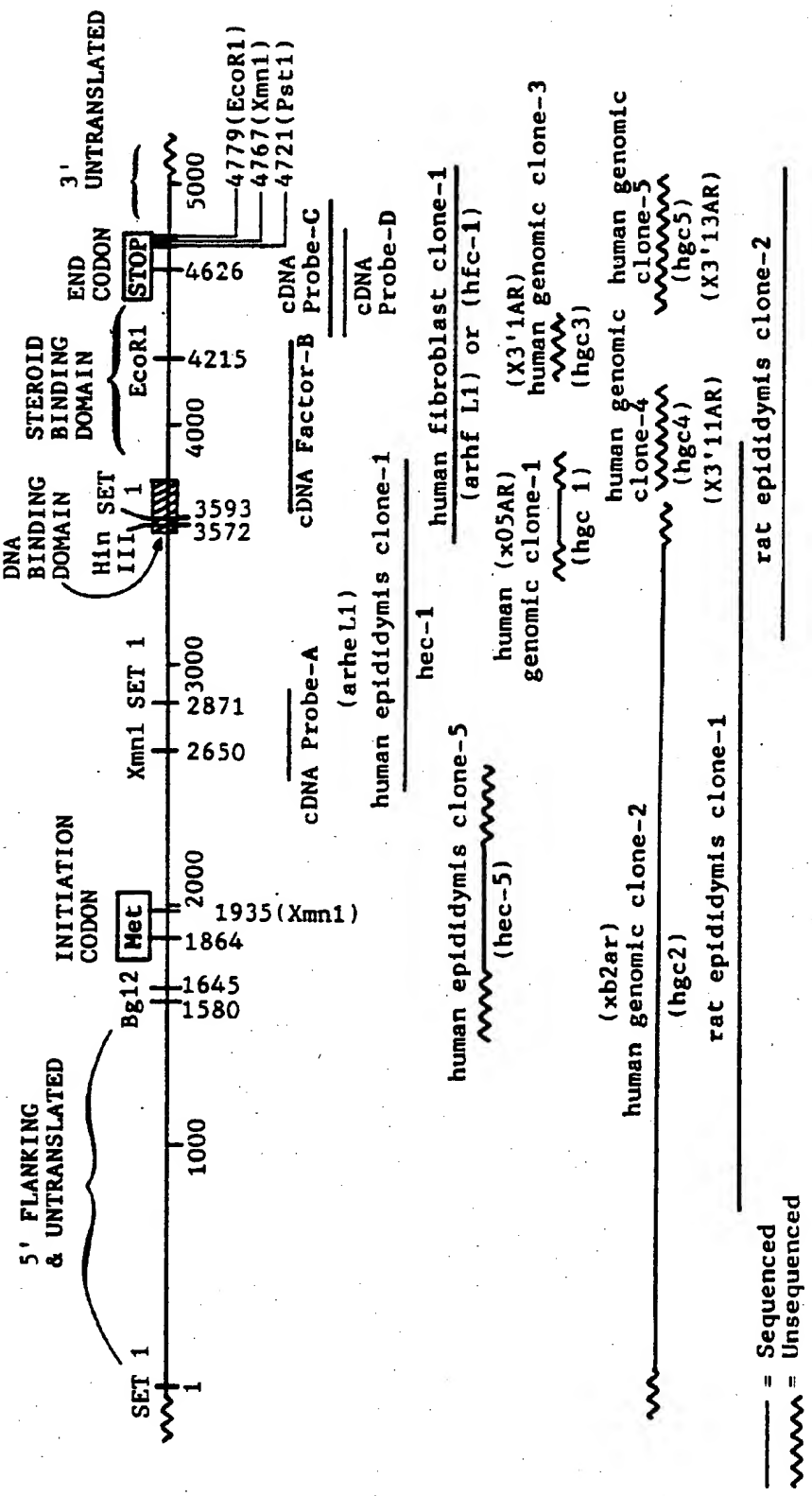


FIGURE 2 (Page 2 of 2)

FIG. 3

COMPILED CLONE MAP OF THE HUMAN ANDROGEN RECEPTOR



```
1     TAATAACTCAGTTCTTATTTGCACCTACTTCAGTGGACACTGAATTTGGAAGGTGGAGGATTTTGTTTTTTTCTTTTAAGATCTGGGCATCTTTTGAATCTACCCTTCAAGTATTAAGA
120   GACAGACTGTGAGCCTAGCAGGGCAGATCTTGTCCACCGTGTGTCTTCTTCTGCACGAGACTTTGAGGCTGTCAGAGCGCTTTTTGCGTGGTTGCTCCCGCAAGTTTCCTTCTCTGGAGC
240   TTCCCGCAGGTGGGCAGCTAGCTGCAGCGACTACCGCATCATCACAGCCTGTTGAACTCTTCTGAGCAAGAGAAGGGGAGGCGGGGTAAGGGAAGTAGGTGGAAGATTCAGCCAAGCTCA
                  ===

                                    10                  20                  30
360   AGGATGGAAGTGCAGTTAGGGCTGGGAAGGGTCTACCCTCGGCCGCCGTCCAAGACCTACCGAGGAGCTTTCCAGAATCTGTTCCAGAGCGTGCGCGAAGTGATCCAGAACCCGGGCCCC
        MetGluValGlnLeuGlyLeuGlyArgValTyrProArgProProSerLysThrTyrArgGlyAlaPheGlnAsnLeuPheGlnSerValArgGluValIleGlnAsnProGlyPro
       40                  50                  60                  70
480   AGGCACCCAGAGGCCGCGAGCGCAGCACCTCCCGGCGCCAGTTTGCTGCTGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGAG
      ArgHisProGluAlaAlaSerAlaAlaProProGlyAlaSerLeuLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlu
       80                  90                  100                 110
600   ACTAGCCCCAGGCAGCAGCAGCAGCAGCAGGGTGAGGATGGTTCTCCCCAAGCCCATCGTAGAGGCCCCACAGGCTACCTGGTCCTGGATGAGGAACAGCAACCTTCACAGCCGCAGTCG
      ThrSerProArgGlnGlnGlnGlnGlnGlnGlyGluAspGlySerProGlnAlaHisArgArgGlyProThrGlyTyrLeuValLeuAspGluGluGlnGlnProSerGlnProGlnSer
      120                 130                 140                 150
720   GCCCTGGAGTGCCACCCCGAGAGAGGTTGCGTCCCAGAGCCTGGAGCCGCCGTGGCCGCCAGCAAGGGGCTGCCGCAGCAGCTGCCAGCACCTCCGGACGAGGATGACTCAGCTGCCCCA
      AlaLeuGluCysHisProGluArgGlyCysValProGluProGlyAlaAlaValAlaAlaSerLysGlyLeuProGlnGlnLeuProAlaProProAspGluAspAspSerAlaAlaPro
      160                 170                 180                 190
840   TCCACGTTGTCCCTGCTGGGCCCCACTTTCCCCGGCTTAAGCAGCTGCTCCGCTGACCTTAAAGACATCCTGAGCGAGGCCAGCACCATGCAACTCCTTCAGCAACAGCAGCAGGAAGCA
      SerThrLeuSerLeuLeuGlyProThrPheProGlyLeuSerSerCysSerAlaAspLeuLysAspIleLeuSerGluAlaSerThrMetGlnLeuLeuGlnGlnGlnGlnGluAla
      200                 210                 220                 230
960   GTATCCGAAGGCAGCAGCAGCGGGAGAGCGAGGGAGGCCTCGGGGGCTCCCACTTCCTCCAAGGACAATTACTTAGGGGGCACTTCGACCATTTCTGACAACGCCAAGGAGTTGTGTAAG
      ValSerGluGlySerSerSerGlyArgAlaArgGluAlaSerGlyAlaProThrSerSerLysAspAsnTyrLeuGlyGlyThrSerThrIleSerAspAsnAlaLysGluLeuCysLys
      240                 250                 260                 270
1080  GCAGTGTCGGTGTCCATGGGCCTGGGTGTGGAGGCGTTGGAGCATCTGAGTCCAGGGGAACAGCTTCGGGGGGATTGCATGTACGCCCCACTTTTGGGAGTTCCACCCGCTGTGCGTCCC
      AlaValSerValSerMetGlyLeuGlyValGluAlaLeuGluHisLeuSerProGlyGluGlnLeuArgGlyAspCysMetTyrAlaProLeuLeuGlyValProProAlaValArgPro
      280                 290                 300                 310
1200  ACTCCTTGTGCCCCATTGGCCGAATGCAAAGGTTCTCTGCTAGACGACAGCGCAGGCAAGAGCACTGAAGATACTGCTGAGTATTCCCCTTTCAAGGGAGGTTACACCAAAGGGCTAGAA
      ThrProCysAlaProLeuAlaGluCysLysGlySerLeuLeuAspAspSerAlaGlyLysSerThrGluAspThrAlaGluTyrSerProPheLysGlyGlyTyrThrLysGlyLeuGlu
      320                 330                 340                 350
1320  GGCGAGAGCCTAGGCTGCTCTGGCAGCGCTGCAGCAGGGAGCTCCGGGACACTTGAACTGCCGTCTACCCTGTCTCTCTACAAGTCCGGAGCACTGGACGAGGCAGCTGCGTACCAGAGT
      GlyGluSerLeuGlyCysSerGlySerAlaAlaAlaGlySerSerGlyThrLeuGluLeuProSerThrLeuSerLeuTyrLysSerGlyAlaLeuAspGluAlaAlaAlaTyrGlnSer
      360                 370                 380                 390
1440  CGCGACTACTACAACTTTCCACTGGCTCTGGCCGGACCGCCGCCCCCTCCGCCGCCTCCCCATCCCCACGCTCGCATCAAGCTGGAGAACCCGCTGGACTACGGCAGCGCCTGGGCGGCT
      ArgAspTyrTyrAsnPheProLeuAlaLeuAlaGlyProProProProProProProHisProHisAlaArgIleLysLeuGluAsnProLeuAspTyrGlySerAlaTrpAlaAla
      400                 410                 420                 430
1560  GCGGCGGCGCAGTGCCGCTATGGGGACCTGGCGAGCCTGCATGGCGCGGGTGCAGCGGGACCCGGTTCTGGGTCACCCTCAGCCGCCGCTTCCTCATCCTGGCACACTCTCTTCACAGCC
      AlaAlaAlaGlnCysArgTyrGlyAspLeuAlaSerLeuHisGlyAlaGlyAlaAlaGlyProGlySerGlySerProSerAlaAlaAlaSerSerSerTrpHisThrLeuPheThrAla
```


EP 0 365 657 B1


440 450 460 470
1680 GAAGAAGGCCAGTTGTATGGACCGTGTGGTGGTGGTGGGGGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGAGGCGGGAGCTGTAGCCCCC
GluGluGlyGlnLeuTyrGlyProCysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGluAlaGlyAlaValAlaPro
480 490 500 510
1800 TACGGCTACACTCGGCCCCCTCAGGGGCTGGCGGGCCAGGAAAGCGACTTCACCGCACCTGATGTGTGGTACCCTGGCGGCATGGTGAGCAGAGTGCCCTATCCCAGTCCCACTTGTGTC
TyrGlyTyrThrArgProProGlnGlyLeuAlaGlyGlnGluSerAspPheThrAlaProAspValTrpTyrProGlyGlyMetValSerArgValProTyrProSerProThrCysVal
520 530 540 550 *
1920 AAAAGCGAAATGGGCCCCTGGATGGATAGCTACTCCGGACCTTACGGGGACATGCGTTTGGAGACTGCCAGGGACCATGTTTTGCCCATTGACTATTACTTTCCACCCCAGAAGACCTGC
LysSerGluMetGlyProTrpMetAspSerTyrSerGlyProTyrGlyAspMetArgLeuGluThrAlaArgAspHisValLeuProIleAspTyrTyrPheProProGlnLysThrCys
560 * 570 * * 580 590 *
2040 CTGATCTGTGGAGATGAAGCTTCTGGGTGTCACTATGGAGCTCTCACATGTGGAAGCTGCAAGGTCTTCTTCAAAAGAGCCGCTGAAGGGAAACAGAAGTACCTGTGCGCCAGCAGAAAT
LeuIleCysGlyAspGluAlaSerGlyCysHisTyrGlyAlaLeuThrCysGlySerCysLysValPhePheLysArgAlaAlaGluGlyLysGlnLysTyrLeuCysAlaSerArgAsn
600 * 610 * * * 620 630
2160 GATTGCACTATTGATAAATTCCGAAGGAAAAATTGTCCATCTTGTCGTCTTCGGAAATGTTATGAAGCAGGGATGACTCTGGGAGCCCGGAAGCTGAAGAAACTTGGTAATCTGAAACTA
AspCysThrIleAspLysPheArgArgLysAsnCysProSerCysArgLeuArgLysCysTyrGluAlaGlyMetThrLeuGlyAlaArgLysLeuLysLysLeuGlyAsnLeuLysLeu
640 650 660 670
2280 CAGGAGGAAGGAGAGGCTTCCAGCACCACCAGCCCCACTGAGGAGACAACCCAGAAGCTGACAGTGTCACACATTGAAGGCTATGAATGTCAGCCCATCTTTCTGAATGTCCTGGAAGCC
GlnGluGluGlyGluAlaSerSerThrThrSerProThrGluGluThrThrGlnLysLeuThrValSerHisIleGluGlyTyrGluCysGlnProIlePheLeuAsnValLeuGluAla
680 690 700 710
2400 ATTGAGCCAGGTGTAGTGTGTGCTGGACACGACAACAACCAGCCCGACTCCTTTGCAGCCTTGCTCTCTAGCCTCAATGAACTGGGAGAGAGACAGCTTGTACACGTGGTCAAGTGGGCC
IleGluProGlyValValCysAlaGlyHisAspAsnAsnGlnProAspSerPheAlaAlaLeuLeuSerSerLeuAsnGluLeuGlyGluArgGlnLeuValHisValValLysTrpAla
720 730 740 750
2520 AAGGCCTTGCCTGGCTTCCGCAACTTACACGTGGACGACCAGATGGCTGTCATTCAGTACTCCTGGATGGGGCTCATGGTGTTTGCCATGGGCTGGCGATCCTTCACCAATGTCAACTCC
LysAlaLeuProGlyPheArgAsnLeuHisValAspAspGlnMetAlaValIleGlnTyrSerTrpMetGlyLeuMetValPheAlaMetGlyTrpArgSerPheThrAsnValAsnSer
760 770 780 790
2640 AGGATGCTCTACTTCGCCCCTGATCTGGTTTTCAATGAGTACCGCATGCACAAGTCCCGGATGTACAGCCAGTGTGTCCGAATGAGGCACCTCTCTCAAGAGTTTGGATGGCTCCAAATC
ArgMetLeuTyrPheAlaProAspLeuValPheAsnGluTyrArgMetHisLysSerArgMetTyrSerGlnCysValArgMetArgHisLeuSerGlnGluPheGlyTrpLeuGlnIle
800 810 820 830
2760 ACCCCCCAGGAATTCCTGTGCATGAAAGCACTGCTACTCTTCAGCATTATTCCAGTGGATGGGCTGAAAAATCAAAAATTCTTTGATGAACTTCGAATGAACTACATCAAGGAACTCGAT
ThrProGlnGluPheLeuCysMetLysAlaLeuLeuLeuPheSerIleIleProValAspGlyLeuLysAsnGlnLysPhePheAspGluLeuArgMetAsnTyrIleLysGluLeuAsp
840 850 860 870
2880 CGTATCATTGCATGCAAAAGAAAAAATCCCACATCCTGCTCAAGACGCTTCTACCAGCTCACCAAGCTCCTGGACTCCGTGCAGCCTATTGCGAGAGAGCTGCATCAGTTCACTTTTGAC
ArgIleIleAlaCysLysArgLysAsnProThrSerCysSerArgArgPheTyrGlnLeuThrLysLeuLeuAspSerValGlnProIleAlaArgGluLeuHisGlnPheThrPheAsp
880 890 900 910
3000 CTGCTAATCAAGTCACACATGGTGAGCGTGGACTTTCCGGAAATGATGGCAGAGATCATCTCTGTGCAAGTGCCCAAGATCCTTTCTGGGAAAGTCAAGCCCATCTATTTCCACACCCAG
LeuLeuIleLysSerHisMetValSerValAspPheProGluMetMetAlaGluIleIleSerValGlnValProLysIleLeuSerGlyLysValLysProIleTyrPheHisThrGln

3120 TGAAGCATTGGAAACCCTATTTCCCCACCCCAGCTCATGCCCCCTTTCAGATGTCTTCTGCCTGTTATAACTCTGCACTACTCCTCTGCAGTGCCTTGGGGAATTTCCTCTATTGATGTA
End
3240 CAGTCTGTCATGAACATGTTCCTGAATTCTATTTGCTGGGCTTTTTTTTTCTCTTTCTCTCCTTTCTTTTTCTTCTTCCCTCCCTATCTAACCCTCCCATGGCACCTTCAGACTTTGCTT
3360 CCCATTGTGGCTCCTATCTGTGTTTTGAATGGTGTTGTATGCCTTTAAATCTGTGATGATCCTCATATGGCCCAGTGTCAAGTTGTGCTTGTTTACAGCACTACTCTGTGCCAGCCACAC
3480 AAACGTTTACTTATCTTATGCCACGGGAAGTTTAGAGAGCTAAGATTATCTGGGGAAATCAAAACAAAAAACAAGCAAACAAAAAAAAAA

FIGURE 5 (Page 1 of 3)

```
ATCCCTAGGAGCCAGCCTGCTGGGAGAACCAGAGGGTCCGGAGCAAACCTGGAGGCTGAGAGGGCATCAGAGGGGAAAAGACTGAGTTAGCCACTCCAGTGCCATACAGAAGCTTAAGGG   120
ACATACCACGCCAGCCCCAGCCCAGCGACAGCCAACGCCTGTTGCAGAGCGGCGGCTTCGAAGCCGCCGCCCAGAAGCTGCCCTTTCCTCTTCGGTGAAGTTTCTAAAAGCTGCGGGAGA   240
CTCGGAGGAAGCGAAGAAAGTGTCCGGTAGGACTACGACTGCCTTTGTCCTCCTCCCTCCTACCCCTACCCCTCCTGGGTCCCCTCTCCCTGAGCGGACTAGGCAGGCTTCCTGGCCAGC   360
CCTCTCCCCTACACCACCAGCTCTGCCAGCCAGTTTGCACAGAGGTAACTCCCTTTGGCTGAAAGCAGACGAGCTTGTTGCCCATTGGAAGGGAGGCTTTTGGGAGCCCAGAGACTGAGG   480
AGCAACAGCACGCTGGAGAGTCCCTGATTCCAGGTTCTCCCCCCTGCACCTCCTACTGCCCGCCCCTCACCCTGTGTGTGCAGCTAGAATTGAAAAGATGAAAAGACAGTTGGGGCTTCA   600
GTAGTCGAAAGCAAAACAAAAGCAAAAAGAAAACAAAAAGAAAATAGCCCAGTTCTTATTTGCACCTGCTTCAGTGGACATTGACTTTGGAAGGCAGAGAATTTTCCTTCCCCCCAGTCA   720
AGCTTTGAGCATCTTTTAATCTGTTCTTCAAGTATTTAGGGACAAACTGTGAAACTAGCAGGGCAGATCCTGTCTAGCGCGTGCCTTCCTTTACAGGAGACTTTGAGGCTATCTGGGCGC   840
TCCCCCCCCTCCCTGCAAGTTTTCTTCCCTGGAGCTTCCCGCAGGTGGGCAGCTAGCTGCAGATACTACATCATCAGTCAGTAGAACTCTTCAGAGCAAGAGACGAGGAGGCAGGATAAG   960

                                                                    10                  20
GGAATTCGGTGGAAGCTAGAGACAAGCTAAAGGATGGAGGTGCAGTTAGGGCTGGGAAGGGTCTACCCACGGCCCCCGTCCAAGACCTATCGAGGAGCGTTCCAGAATCTGTTCCAGAGC  1080
                                    MetGluValGlnLeuGlyLeuGlyArgValTyrProArgProProSerLysThrTyrArgGlyAlaPheGlnAsnLeuPheGlnSer
 30                  40                  50                  60
GTGCGCGAAGCGATCCAGAACCCGGGCCCCAGGCACCCTGAGGCCGCTAGCATAGCACCTCCCGGTGCCTGTTTACAGCAGCGGCAGGAGACTAGCCCCCGGCGGCGGCGGCGGCAGCAG  1200
ValArgGluAlaIleGlnAsnProGlyProArgHisProGluAlaAlaSerIleAlaProProGlyAlaCysLeuGlnGlnArgGlnGluThrSerProArgArgArgArgArgGlnGln
 70                  80                  90                  100
CACCCTGAGGATGGCTCTCCTCAAGCCCACATCAGAGGCACCACAGGCTACCTGGCCCTGGAGGAGGAACAGCAGCCTTCACAGCAGCAGTCAGCCTCCGAGGGCCACCCTGAGAGCGGC  1320
HisProGluAspGlySerProGlnAlaHisIleArgGlyThrThrGlyTyrLeuAlaLeuGluGluGluGlnGlnProSerGlnGlnGlnSerAlaSerGluGlyHisProGluSerGly
110                 120                 130                 140
TGCCTCCCGGAGCCTGGAGCTGCCACGGCTCCTGGCAAGGGGCTGCCGCAGCAGCCACCAGCTCCTCCAGATCAGGATGACTCAGCTGCCCCATCCACGTTGTCCCTACTGGGCCCCACT  1440
CysLeuProGluProGlyAlaAlaThrAlaProGlyLysGlyLeuProGlnGlnProProAlaProProAspGlnAspAspSerAlaAlaProSerThrLeuSerLeuLeuGlyProThr
150                 160                 170                 180
TTCCCAGGCTTAAGCAGCTGCTCCGCAGACATTAAAGACATCCTGAGCGAGGCCGGCACCATGCAACTTCTTCAGCAGCAGCAGCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAG  1560
PheProGlyLeuSerSerCysSerAlaAspIleLysAspIleLeuSerGluAlaGlyThrMetGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln
190                 200                 210                 220
CAGCAGCAACAGCAGCAGGAGGTAATATCCGAAGGCAGCAGCAGCGTGAGAGCAAGGGAGGCCACTGGGGCTCCCTCTTCCTCCAAGGATAGTTACCTAGGGGGCAATTCGACCATATCT  1680
GlnGlnGlnGlnGlnGlnGluValIleSerGluGlySerSerSerValArgAlaArgGluAlaThrGlyAlaProSerSerSerLysAspSerTyrLeuGlyGlyAsnSerThrIleSer
230                 240                 250                 260
GACAGTGCCAAGGAGTTGTGTAAAGCAGTGTCTGTGTCCATGGGGTTGGGTGTGGAAGCACTGGAACATCTGAGTCCAGGGGAGCAGCTTCGGGGCGACTGCATGTACGCGTCGCTCCTG  1800
AspSerAlaLysGluLeuCysLysAlaValSerValSerMetGlyLeuGlyValGluAlaLeuGluHisLeuSerProGlyGluGlnLeuArgGlyAspCysMetTyrAlaSerLeuLeu
270                 280                 290                 300
GGAGGTCCACCCGCCGTGCGTCCCACTCCTTGTGCGCCTCTGGCCGAATGCAAAGGTCTTTCCCTGGACGAAGGCCCGGGCAAAGGCACTGAAGAGACTGCTGAGTATTCCTCTTTCAAG  1920
GlyGlyProProAlaValArgProThrProCysAlaProLeuAlaGluCysLysGlyLeuSerLeuAspGluGlyProGlyLysGlyThrGluGluThrAlaGluTyrSerSerPheLys
310                 320                 330                 340
GGAGGTTACGCCAAAGGGTTGGAAGGTGAGAGTCTGGGCTGCTCTGGCAGCAGTGAAGCAGGTAGCTCTGGGACACTTGAGATCCCGTCCTCACTGTCTCTGTATAAGTCTGGAGCAGTA  2040
GlyGlyTyrAlaLysGlyLeuGluGlyGluSerLeuGlyCysSerGlySerSerGluAlaGlySerSerGlyThrLeuGluIleProSerSerLeuSerLeuTyrLysSerGlyAlaVal
```

FIGURE 5 (Page 2 of 3)

```
350                       360                       370                       380
GACGAGGCAGCAGCATACCAGAATCGCGACTACTACAACTTTCCGCTCGCTCTGTCCGGGCCGCCGCACCCCCCGCCCCCTACCCATCCACACGCCCGCATCAAGCTGGAGAACCCGTCG 2160
AspGluAlaAlaAlaTyrGlnAsnArgAspTyrTyrAsnPheProLeuAlaLeuSerGlyProProHisProProProProThrHisProHisAlaArgIleLysLeuGluAsnProSer
390                       400                       410                       420
GACTACGGCAGCGCCTGGGCTGCGGCGGCAGCGCAATGCCGCTATGGGGACTTGGCTAGCCTACATGGAGGGAGTGTAGCCGGACCCAGCACTGGATCGCCCCCAGCCACCGCCTCTTCT 2280
AspTyrGlySerAlaTrpAlaAlaAlaAlaAlaGlnCysArgTyrGlyAspLeuAlaSerLeuHisGlyGlySerValAlaGlyProSerThrGlySerProProAlaThrAlaSerSer
430                       440                       450                       460
TCCTGGCATACTCTCTTCACAGCTGAAGAAGGCCAATTATATGGGCCAGGAGGCGGGGGCGGCAGCAGTAGCCCAAGCGATGCTGGGCCTGTAGCCCCCTATGGCTACACTCGGCCCCCT 2400
SerTrpHisThrLeuPheThrAlaGluGluGlyGlnLeuTyrGlyProGlyGlyGlyGlyGlySerSerSerProSerAspAlaGlyProValAlaProTyrGlyTyrThrArgProPro
470                       480                       490                       500
CAGGGGCTGGCAAGCCAGGAGGGTGACTTCTCTGCCTCTGAAGTGTGGTATCCTGGTGGAGTTGTGAACAGAGTCCCCTATCCCAGTCCCAGTTGTGTTAAAAGTGAAATGGGACCTTGG 2520
GlnGlyLeuAlaSerGlnGluGlyAspPheSerAlaSerGluValTrpTyrProGlyGlyValValAsnArgValProTyrProSerProSerCysValLysSerGluMetGlyProTrp
510                       520                       530                       540
ATGGAGAACTACTCCGGACCTTATGGGGACATGCGTTTGGACAGTACCAGGGACCACGTTTTACCCATCGACTATTACTTCCCACCCCAGAAGACCTGCCTGATCTGTGGAGATGAAGCT 2640
MetGluAsnTyrSerGlyProTyrGlyAspMetArgLeuAspSerThrArgAspHisValLeuProIleAspTyrTyrPheProProGlnLysThrCysLeuIleCysGlyAspGluAla
550                       560                       570                       580
TCTGGTTGTCACTACGGAGCTCTCACTTGTGGCAGCTGCAAGGTCTTCTTCAAAAGAGCTGCGGAAGGGAAACAGAAGTATCTATGTGCCAGCAGAAATGATTGCACCATTGATAAATTT 2760
SerGlyCysHisTyrGlyAlaLeuThrCysGlySerCysLysValPhePheLysArgAlaAlaGluGlyLysGlnLysTyrLeuCysAlaSerArgAsnAspCysThrIleAspLysPhe
590                       600                       610                       620
CGGAGGAAAAATTGTCCATCGTGTCGTCTCCGGAAATGTTATGAAGCAGGGATGACTCTGGGAGCTCGTAAGCTGAAGAAACTTGGAAATCTCAAACTACAGGAAGAAGGAGAAAACTCC 2880
ArgArgLysAsnCysProSerCysArgLeuArgLysCysTyrGluAlaGlyMetThrLeuGlyAlaArgLysLeuLysLysLeuGlyAsnLeuLysLeuGlnGluGluGlyGluAsnSer
630                       640                       650                       660
AGTGCTGGTAGCCCCACTGAGGACCCATCCCAGAAGATGACTGTATCACACATTGAAGGCTATGAATGTCAACCTATCTTTCTTAATGTCCTGGAAGCCATTGAGCCAGGAGTGGTGTGT 3000
SerAlaGlySerProThrGluAspProSerGlnLysMetThrValSerHisIleGluGlyTyrGluCysGlnProIlePheLeuAsnValLeuGluAlaIleGluProGlyValValCys
670                       680                       690                       700
GCCGGACATGACAACAACCAGCCTGATTCCTTTGCTGCCTTGTTATCTAGTCTCAACGAGCTTGGCGAGAGACAGCTTGTACATGTGGTCAAGTGGGCCAAGGCCTTGCCTGGCTTCCGC 3120
AlaGlyHisAspAsnAsnGlnProAspSerPheAlaAlaLeuLeuSerSerLeuAsnGluLeuGlyGluArgGlnLeuValHisValValLysTrpAlaLysAlaLeuProGlyPheArg
710                       720                       730                       740
AACTTGCATGTGGATGACCAGATGGCAGTCATTCAGTATTCCTGGATGGGACTGATGGTATTTGCCATGGGTTGGCGGTCCTTCACTAATGTCAACTCTAGGATGCTCTACTTTGCACCT 3240
AsnLeuHisValAspAspGlnMetAlaValIleGlnTyrSerTrpMetGlyLeuMetValPheAlaMetGlyTrpArgSerPheThrAsnValAsnSerArgMetLeuTyrPheAlaPro
750                       760                       770                       780
GACCTGGTTTTCAATGAGTATCGCATGCACAAGTCTCGAATGTACAGCCAGTGCGTGAGGATGAGGCACCTTTCTCAAGAGTTTGGATGGCTCCAGATAACCCCCCAGGAATTCCTGTGC 3360
AspLeuValPheAsnGluTyrArgMetHisLysSerArgMetTyrSerGlnCysValArgMetArgHisLeuSerGlnGluPheGlyTrpLeuGlnIleThrProGlnGluPheLeuCys
790                       800                       810                       820
ATGAAAGCACTGCTACTCTTCAGCATTATTCCAGTGGATGGGCTGAAAAATCAAAAATTCTTTGATGAACTTCGAATGAACTACATCAAGGAACTTGATCGCATCATTGCATGCAAAAGA 3480
MetLysAlaLeuLeuLeuPheSerIleIleProValAspGlyLeuLysAsnGlnLysPhePheAspGluLeuArgMetAsnTyrIleLysGluLeuAspArgIleIleAlaCysLysArg
```

FIGURE 5 (page 3 of 3)

```
830                       840                       850                       860
AAAAATCCCACATCCTGCTCAAGGCGCTTCTACCAGCTCACCAAGCTCCTGGATTCTGTGCAGCCTATTGCAAGAGAGCTGCATCAATTCACTTTTGACCTGCTAATCAAGTCCCATATG 3600
LysAsnProThrSerCysSerArgArgPheTyrGlnLeuThrLysLeuLeuAspSerValGlnProIleAlaArgGluLeuHisGlnPheThrPheAspLeuLeuIleLysSerHisMet
870                       880                       890                       900
GTGAGCGTGGACTTTCCTGAAATGATGGCAGAGATCATCTCTGTGCAAGTGCCCAAGATCCTTTCTGGGAAAGTCAAGCCCATCTATTTCCACACACAGTGAAGATTTGGAAACCCTAAT 3720
ValSerValAspPheProGluMetMetAlaGluIleIleSerValGlnValProLysIleLeuSerGlyLysValLysProIleTyrPheHisThrGlnEnd

ACCCAAACCCACCTTGTTCCCTTTTCAGATGTCTTCTGCCTGTTATATAACTCTGCACTACTTCTCTGCATTGCCTTGGGGGAAATTCCTCTACTGATGTACAGTCTGTCATGAACATGT 3840
TCCCCAAGTTCTATTTCCTGGGCTTTTCCTTCTTTCTTTTTCTTCTTCTCTGCCTCTTTTACCCTCCCATGGCACATTTTGAATCCGCTGCGTGTTGTGGCTCCTGCCTGTGTTTTGAGT 3960
TTTGTTGTATTTCTTCAAGTCTGTGATGATCTTCTTGTGGCCCAGTGTCAACTGTGCTTGTTTATAGCACTGTGCTGTGTGCCAACCAAGCAAATGTTTACTCACCTTATGCCATGGCAA 4080
GTTTAGAGAGCTATAAGTATCTTGGGAAGAAACAAACAGAGAGAGTAAAAAAACC(A)45
```

FIGURE 6